When is an input state always better than the others ?: universally optimal input states for statistical inference of quantum channels

Keiji Matsumoto

National Institute of Informatics, HItotsubashi 2-1-2, Chiyoda-ku, Tokyo 101-8430, e-mail: keiji@nii.ac.jp

September 17, 2012

Abstract

Statistical estimation and test of unknown channels have attracted interest of many researchers. In optimizing the process of inference, an important step is optimization of the input state, which in general do depend on the kind of inference (estimation or test, etc.), on the error measure, and so on. But sometimes, there is a universally optimal input state, or an input state best for all the statistical inferences and for all the risk functions. In the paper, the existence of a universally optimal state is shown for group covariant/contravariant channels, unital qubit channels and some measurement families. To prove these results, theory of "comparison of state families" are used. We also discuss about effectiveness of entanglement and adaptation of input states.

1   Introduction

Statistical estimation and test of unknown channels have attracted interests of many researchers. Below, let $\{\Lambda_\theta\}_{\theta\in\Theta}$ be a family of unknown channels, where $\theta\in\Theta$ is the unknown parameter. In optimizing the process of inference, one has to optimize not only the measurement performed upon the output state $\Lambda_\theta\otimes\mathbf{I}(\rho_{in})$, but also the input state ρ_{in}. (One may also use a process POVM [15], operators $\{M_t\}_{t\in\mathcal{D}}$ such that $\sum_{t\in\mathcal{D}}M_t=\mathbf{1}\otimes\mathrm{tr}_{\mathcal{H}_R}\rho_{in}^T$. But then one also has to optimize $\mathrm{tr}_{\mathcal{H}_R}\rho_{in}$. Since an optimal input state ρ_{in} is a pure state, optimization of $\mathrm{tr}_{\mathcal{H}_R}\rho_{in}$ is equivalent to optimization of ρ_{in} .)

In general, optimal input states depend on whether we are estimating state or testing hypothesis about unknown channels; they also depend on error measure, and detail of the setting (Bayesian, minimax, unbiased estimation, Neyman-Pearson test, etc.).

In some cases, however, the situation is less complicated. For example, [3] deals with estimation of group transform $\{U_g\}_{g\in\mathcal{G}}$, where $g\to U_g$ is a represen-

tation of the group $\mathcal{G}$ and g is unknown and to be estimated. They had shown that there is an input state which is optimal with respect to any $\mathcal{G}$-invariant loss functions. (In case of $\mathcal{G} = \mathrm{SU}\,(d)$ and $U_g = g$, maximally entangled states between the input space and the auxiliary space are optimal.) Meantime, [4] treats estimation of $\mathrm{SU}\,(2)$ channel by an unbiased estimator, and 'the loss function' here is the mean square error matrix of the estimate $\hat{\theta}$ of the unknown real vector θ which parameterizes $\mathcal{G} = \mathrm{SU}\,(2)$. Since the space of matrices is not totally ordered, the existence of the minimum is non-trivial. Put differently, if the loss is scalar valued increasing function of a mean square error matrix, then, maximally entangled states are optimal. Also, [11] studies discrimination of a pair of generalized Pauli matrices, and shows maximally entangled states minimize Bayesian error probability for any prior distributions. In case of qubits, they extended their result to minimax error probability [13]. Another example of such study is [16], where discrimination of two unitary operation is discussed. They found that minimizers of Bayesian error probability and the error probability of unambiguous discrimination are the same.

These results motivate the following definition: we say the input is *universally optimal* for the family $\{\Lambda_\theta\}_{\theta \in \Theta}$, roughly speaking, if it is optimal for all the statistical inferences and for all the loss functions. (The rigorous definition will be given later.) We show that a universally optimal state exists (not necessarily uniquely) in case of group covariant and contravariant channels, unital qubit channels and some measurement families.

To prove these results, we have recourse to the theory of "comparison of state families" [2][10]; we write $\{\rho_\theta\}_{\theta \in \Theta} \succeq^c \{\sigma_\theta\}_{\theta \in \Theta}$ if the family $\{\rho_\theta\}_{\theta \in \Theta}$ is more informative than another family $\{\sigma_\theta\}_{\theta \in \Theta}$ with respect to any kind of statistical inferences. Then, our target is to prove

$$\forall \rho' \ \{(\Lambda_\theta \otimes \mathbf{I})\,(\rho_{\mathrm{opt}})\}_{\theta \in \Theta} \succeq_c \{(\Lambda_\theta \otimes \mathbf{I})\,(\rho')\}_{\theta \in \Theta}, \ \forall \rho'$$

for an input ρ_{opt}. In particular, we utilize sufficient conditions for $\{\rho_\theta\} \succeq^c \{\sigma_\theta\}$, Proposition 2.2 and Lemma 2.4.

Based on these results, some related topics are discussed. The first topic is effect of entanglement between the input space and the auxiliary space. For example, in [11][12][13], they study the condition that Bayes risk and minimax risk of discrimination of two unital qubit channels is smaller on an entangled state than on any separable state. In our case, in Sections 4-6 it is shown that a maximally entangle is universally optimal for some channel families. But there might be a separable state which is as good as maximally entangled states. So we question whether the entanglement is really needed or not.

The second topic discussed is the existence of universally optimal states under the setting where the given channel can be used for several times.

The paper is organized as follows. In Section 2, besides introducing notations and definitions, the theory of comparison of state families is explained. In Sections 3, 4 and 5, universally optimal input states are established for a pair of unitary operations, covariant/contravariant channel families, and unital qubit channel families, respectively. In the proof, Proposition 2.2 is used. In Section 6,

with the help of Lemma 2.4, we investigate universally optimal states for some families which consist of a pair of measurements. In Section 7, the family of SU (d) is studied. In $d = 2$-case, it is shown, with recourse to Theorem 5.3 in Section 5, that maximally entangled states are universally optimal. On the other hand, in $d \geq 3$-case, it is shown that any state is optimal for some statistical inferences. In Section 8, we investigate the conditions that an entangled state is strictly universally better than any separable states. In Section 9, universally optimal input states in case of iterative use of the given channel is studied.

2 Preliminaries

2.1 Settings, conventions and notations

Below, $\mathcal{H}_{in}$, $\mathcal{H}_{out}$, $\mathcal{H}_R$ etc. are finite dimensional Hilbert spaces, and $\mathcal{B}(\mathcal{H}_{in})$, for example, are the set of linear operators over $\mathcal{H}_{in}$. $\mathbf{1}_{in}$ and $\mathbf{I}_{in}$ denotes identity operator over $\mathcal{H}_{in}$ and over $\mathcal{B}(\mathcal{H}_{in})$, respectively. A channel is a trace preserving completely positive (CPTP, hereafter) map from $\mathcal{B}(\mathcal{H}_{in})$ to $\mathcal{B}(\mathcal{H}_{out})$, and is represented by Λ, Υ , etc. with subscripts and superscripts.

To do some statistical inference about a family $\{\Lambda_\theta\}_{\theta \in \Theta}$ of channels $\Lambda_\theta :$ $\mathcal{B}(\mathcal{H}_{in}) \to \mathcal{B}(\mathcal{H}_{out})$, a statistician prepares an input state $\rho_{in} \in \mathcal{B}(\mathcal{H}_{in} \otimes \mathcal{H}_R)$, sends its $\mathcal{H}_{in}$-part to Λ_θ, obtaining $\Lambda_\theta \otimes \mathbf{I}_R (\rho_{in})$ as the output. To the output $\Lambda_\theta \otimes \mathbf{I}_R (\rho_{in})$, the statistician apply a measurement with POVM M which takes values in decision space $\mathcal{D}$ (an element of $\mathcal{D}$ is usually denoted by t). Without loss of generality, throughout the paper, we suppose ρ_{in} is pure, and thus we suppose $\dim \mathcal{H}_{in} = \dim \mathcal{H}_R = d$. For a state vector $|\psi\rangle \in \mathcal{H}_{in} \otimes \mathcal{H}_R$,

$$\rho_\psi := \mathrm{tr}_{\mathcal{H}_R} |\psi\rangle \langle\psi| \in \mathcal{B}(\mathcal{H}_{in}). \tag{1}$$

A system of vectors $\{|i\rangle\}_{i=1}^d$ is an orthonormal complete basis of $\mathcal{H}_{in}$. Abusing the notation, the same symbol is also used to denote an orhonormal basis of $\mathcal{H}_R$.

$$|\Phi_d\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^d |i\rangle |i\rangle$$

is a maximally entangled state living in $\mathcal{H}_{in} \otimes \mathcal{H}_R$.

Given a linear map Γ from $\mathcal{B}(\mathcal{H}_{in})$ to $\mathcal{B}(\mathcal{H}_{out})$, its Choi-Jamilokovski's representation $Ch(\Gamma)$ is defined by

$$Ch(\Gamma) := \sum_{i,j=1}^d \Gamma(|i\rangle \langle j|) \otimes |i\rangle \langle j|.$$

We also use the following notation:

$$\Upsilon_C(\rho) := C\rho C^\dagger.$$

Given a state ρ and a POVM M, denote $P_\rho^M(B) := \mathrm{tr}\, \rho M(B)$.

When $\Theta \subset \mathcal{D} = \mathbb{R}^m$, we write

$$\mathrm{E}\left[M, \rho_\theta\right] := \int_{t \in \mathcal{D}} t \, \mathrm{d}P_{\rho_\theta}^M(t),$$

$$\mathrm{V}\left[M, \rho_\theta\right] := \left[\int \left(t^i - \theta^i\right)\left(t^j - \theta^j\right) \mathrm{d}P_{\rho_\theta}^M(t)\right].$$

2.2 Comparison of state families

In comparison of input states, we have recourse to the theory of comparison of state families[2][10]. Consider a family $\{\rho_\theta\}_{\theta \in \Theta}$ of states over $\mathcal{H}$ and a family $\{\sigma_\theta\}_{\theta \in \Theta}$ of states over $\mathcal{H}'$. We say $\{\rho_\theta\}_{\theta \in \Theta}$ is *sufficient* to $\{\sigma_\theta\}_{\theta \in \Theta}$ with respect to classical decision problems, and write $\{\rho_\theta\}_{\theta \in \Theta} \succeq^c \{\sigma_\theta\}_{\theta \in \Theta}$, if and only if, for any decision space $\mathcal{D}$ equipped with σ-field $\mathfrak{A}$, any σ-field $\mathfrak{B}$ over Θ, any loss function $l : \Theta \times \mathcal{D} \to \mathbb{R}_+$ which is jointly measurable, any probability measure π over $(\Theta, \mathfrak{B})$, and for any measurement M' over $(\mathcal{D}, \mathfrak{A})$ in $\mathcal{H}'$, there is a measurement M over $(\mathcal{D}, \mathfrak{A})$ in $\mathcal{H}$ such that

$$\int_{\Theta \times \mathcal{D}} l_\theta(t) \, \mathrm{d}P_{\rho_\theta}^M(t) \, \mathrm{d}\pi(\theta) \le \int_{\Theta \times \mathcal{D}} l_\theta(t) \, \mathrm{d}P_{\sigma_\theta}^{M'}(t) \, \mathrm{d}\pi(\theta).$$

When $\{\rho_\theta\}_{\theta \in \Theta} \succeq^c \{\sigma_\theta\}_{\theta \in \Theta}$ and $\{\sigma_\theta\}_{\theta \in \Theta} \succeq^c \{\rho_\theta\}_{\theta \in \Theta}$ holds, we write $\{\rho_\theta\}_{\theta \in \Theta} \equiv^c \{\sigma_\theta\}_{\theta \in \Theta}$.

Lemma 2.1 $\{\rho_\theta\}_{\theta \in \Theta} \succeq^c \{\sigma_\theta\}_{\theta \in \Theta}$ *holds if and only if, for any measurement M on $(\mathcal{D}, \mathfrak{A})$, there is a measurement M' on $(\mathcal{D}, \mathfrak{A})$ such that $P_{\rho_\theta}^{M'} = P_{\sigma_\theta}^M$.*

Due to Lemma 2.1, $\{\rho_\theta\}_{\theta \in \Theta} \succeq^c \{\sigma_\theta\}_{\theta \in \Theta}$ has very strong implications: whatever the settings are, and whatever the error measures are chosen, $\{\rho_\theta\}_{\theta \in \Theta}$ is always better than $\{\sigma_\theta\}_{\theta \in \Theta}$. For example, for any decision space $\mathcal{D}$ equipped with σ-field $\mathfrak{A}$, any loss function $l : \Theta \times \mathcal{D} \to \mathbb{R}_+$ such that $l_\theta(\cdot)$ is measurable, the minimax risk is always smaller on $\{\rho_\theta\}_{\theta \in \Theta}$ than on $\{\sigma_\theta\}_{\theta \in \Theta}$:

$$\inf_M \sup_{\theta \in \Theta} \int_{\mathcal{D}} l_\theta(t) \, \mathrm{d}P_{\rho_\theta}^M(t) \le \inf_M \sup_{\theta \in \Theta} \int_{\mathcal{D}} l_\theta(t) \, \mathrm{d}P_{\sigma_\theta}^M(t).$$

Also, in hypothesis testing of Neyman-Pearson type, the second error probability of the optimal level α test is also smaller on $\{\rho_\theta\}_{\theta \in \Theta}$ than on $\{\sigma_\theta\}_{\theta \in \Theta}$. That is, letting $\mathcal{D} := \{0, 1\}$, $\Theta_0 \cup \Theta_1 = \Theta$, and

$$l_\theta^{\mathrm{T}}(t) := \left\{ \begin{array}{ll} 1, & (\theta \in \Theta_0 \text{ and } t = 1, \text{ or } \theta \in \Theta_1 \text{ and } t = 0) \\ 0, & \text{otherwise} \end{array} \right.,$$

we have

$$\inf_M \left\{ \int l_1^{\mathrm{T}}(t) \, \mathrm{d}P_{\rho_\theta}^M(t) \, ; \int l_0^{NP}(t) \, \mathrm{d}P_{\rho_\theta}^M(t) \le \alpha \right\}$$

$$\le \inf_M \left\{ \int l_1^{\mathrm{T}}(t) \, \mathrm{d}P_{\sigma_\theta}^M(t) \, ; \int l_0^{NP}(t) \, \mathrm{d}P_{\sigma_\theta}^M(t) \le \alpha \right\}.$$

4

Another example would be unambiguous discrimination: letting $\mathcal{D} := \{0, 1, 2\}$, $\Theta_0 \cup \Theta_1 = \Theta$,

$$l_\theta^{\mathrm{IT}}(t) := \left\{ \begin{array}{ll} \infty, & (\theta \in \Theta_0 \text{ and } t = 1, \text{ or } \theta \in \Theta_1 \text{ and } t = 0) \\ 1, & (\theta \in \Theta_0 \text{ and } t = 2, \text{ or } \theta \in \Theta_1 \text{ and } t = 2) \\ 0, & (\theta \in \Theta_0 \text{ and } t = 0, \text{ or } \theta \in \Theta_1 \text{ and } t = 1) \end{array} \right. ,$$

we have

$$\inf_M \int_{\Theta \times \mathcal{D}} l_\theta^{\mathrm{IT}}(t) \, \mathrm{d}P_{\rho_\theta}^M(t) \, \mathrm{d}\pi(\theta) \leq \inf_M \int_{\Theta \times \mathcal{D}} l_\theta^{\mathrm{IT}}(t) \, \mathrm{d}P_{\sigma_\theta}^M(t) \, \mathrm{d}\pi(\theta).$$

Lastly, let $\Theta \subset \mathcal{D} = \mathbb{R}^m$. Then, mean square error of an unbiased estimator is always better on $\{\rho_\theta\}_{\theta \in \Theta}$ than on $\{\sigma_\theta\}_{\theta \in \Theta}$. That is, for any measurement M' with

$$\mathrm{E}\left[M', \sigma_\theta\right] = \theta,$$

there is a measurement M such that

$$\mathrm{E}\left[M, \rho_\theta\right] = \theta,$$
$$\mathrm{V}\left[M, \rho_\theta\right] = \mathrm{V}\left[M', \sigma_\theta\right].$$

Proposition 2.2 *[10]If there is a trace preserving positive map Γ such that $\Gamma(\rho_\theta) = \sigma_\theta$, we have $\{\rho_\theta\}_{\theta \in \Theta} \succeq^c \{\sigma_\theta\}_{\theta \in \Theta}$.*

Lemma 2.3 *[1]There is a CPTP map Γ with*

$$\Gamma\left(|\psi_+\rangle\langle\psi_+|\right) = |\varphi_+\rangle\langle\varphi_+|, \, \Gamma\left(|\psi_-\rangle\langle\psi_-|\right) = |\varphi_-\rangle\langle\varphi_-|,$$

If and only if

$$|\langle\psi_+|\psi_-\rangle| \leq |\langle\varphi_+|\varphi_-\rangle|.$$

Lemma 2.4 *[10]Suppose $\Theta = \{+, -\}$. If $\{\rho_\theta\}_{\theta \in \Theta} \succeq^c \{\sigma_\theta\}_{\theta \in \Theta}$, then*

$$\|\rho_+ - s\,\rho_-\|_1 \geq \|\sigma_+ - s\,\sigma_-\|_1, \, \forall s \geq 0. \tag{2}$$

If (2) and $[\rho_+, \rho_-] = 0$, then $\{\rho_\theta\}_{\theta \in \Theta} \succeq^c \{\sigma_\theta\}_{\theta \in \Theta}$.

Lemma 2.5 *Suppose $\Theta = \{+, -\}$, $\sigma_\theta \in \mathcal{B}\left(\mathbb{C}^2\right)$, and $[\rho_+, \rho_-] = 0$. If $\{\rho_\theta\}_{\theta \in \Theta} \equiv^c \{\sigma_\theta\}_{\theta \in \Theta}$, we have*

$$[\sigma_+, \sigma_-] = 0.$$

Proof. By definition, $\{\rho_\theta\}_{\theta \in \Theta} \preceq^c \{\sigma_\theta\}_{\theta \in \Theta}$ only if there is a measurement M with

$$\{\rho_\theta\}_{\theta \in \Theta} \preceq^c \left\{P_{\sigma_\theta}^M\right\}_{\theta \in \Theta}.$$

By Lemma 2.4, this is equivalent to

$$\|\rho_+ - s\,\rho_-\|_1 \leq \left\|P_{\sigma_+}^M - s\,P_{\sigma_-}^M\right\|_1, \, \forall s \geq 0.$$

Also, by Lemma 2.4, $\{\rho_\theta\}_{\theta \in \Theta} \equiv^c \{\sigma_\theta\}_{\theta \in \Theta}$ only if

$$\|\rho_+ - s\,\rho_-\|_1 = \|\sigma_+ - s\,\sigma_-\|_1, \ \forall s \geq 0.$$

Therefore, we have

$$\|\sigma_+ - s\,\sigma_-\|_1 \leq \left\|P_{\sigma_+}^M - s\,P_{\sigma_-}^M\right\|_1, \ \forall s \geq 0.$$

Therefore, by the monotonicity of $\|\cdot\|_1$, there is a measurement M such that

$$\|\sigma_+ - s\,\sigma_-\|_1 = \left\|P_{\sigma_+}^M - s\,P_{\sigma_-}^M\right\|_1, \ \forall s \geq 0.$$

Observe the above identity holds if and only if $M = \{M_+, M_-\}$, where M_+ and M_- are the projector onto the positive and the negative eigenvector of $\sigma_+ - s\,\sigma_-$, respectively. Since M does not depends on s, combined with the fact that σ_θ is a qubit state, we have

$$[\sigma_+ - s\,\sigma_-, \sigma_+ - s'\,\sigma_-] = 0,$$

or equivalently, $[\sigma_+, \sigma_-] = 0$. ∎

2.3 Comparison of input states

Consider a family $\{\Lambda_\theta\}_{\theta \in \Theta}$ of channels $\Lambda_\theta : \mathcal{B}(\mathcal{H}_{in}) \to \mathcal{B}(\mathcal{H}_{out})$. We say the input state ρ is *universally better* than ρ' and write $\rho \succeq^c \rho'$ if and only if ρ is better than ρ' for any statistical decision problem on $\{\Lambda_\theta\}_{\theta \in \Theta}$. More formally, $\rho \succeq^c \rho'$ if and only if

$$\{(\Lambda_\theta \otimes \mathbf{I})(\rho)\}_{\theta \in \Theta} \succeq_c \{(\Lambda_\theta \otimes \mathbf{I})(\rho')\}_{\theta \in \Theta}.$$

If $\rho \succeq^c \rho'$ and $\rho' \not\succeq^c \rho$ holds, we say ρ is strictly universally better than ρ', and write $\rho \succ^c \rho'$. If $\rho \succeq^c \rho'$ and $\rho' \succeq^c \rho$ holds, we write $\rho \equiv^c \rho'$ and say that ρ and ρ' are universally equivalent. Obviously,

$$\rho \equiv^c \Lambda_{\mathbf{1} \otimes U}(\rho)$$

for any $U \in \mathrm{SU}(\mathcal{H}_R)$.

Denote

$$R(l, M, \pi, \rho) := \int_{\Theta \times \mathcal{D}} l_\theta(t) \, \mathrm{d}P_{\Lambda_\theta \otimes \mathbf{I}(\rho)}^M(t) \, \mathrm{d}\pi(\theta).$$

An input state $\rho \in \mathcal{B}(\mathcal{H}_{in} \otimes \mathcal{H}_R)$ is said to be *admissible* if and only if, for a decision space $\mathcal{D}$ equipped with a σ-field $\mathfrak{A}$, a σ-field $\mathfrak{B}$ over Θ, a loss function $l : \Theta \times \mathcal{D} \to \mathbb{R}_+$ which is jointly measurable, a probability measure π over $(\Theta, \mathfrak{B})$,

$$\inf_M R(l, M, \pi, \rho) \leq \inf_M R(l, M, \pi, \rho'), \forall \rho. \tag{3}$$

When the inequality in (3) is strict inequality "$<$", ρ is said to be *strictly admissible*.

3 A pair of unitary operations

Let $\Lambda_\theta = \Upsilon_{U_\theta}$, $\Theta = \{+, -\}$ and U_+, $U_- \in \mathrm{SU}(d)$. [16] had discussed discrimination U_+, U_- and computed Bayesian error probability and error probability of unambiguous discrimination. After performing optimization for each case, they found that optimal input states are minimizers of the functional

$$|\psi\rangle \to \left| \langle\psi| U_+^\dagger U_- \otimes \mathbf{1} |\psi\rangle \right|. \tag{4}$$

Indeed, generalizing their result, we can conclude that minimizers of (4) are universally optimal, or optimal for any statistical inference made upon $\{\Lambda_\theta\}_{\theta\in\Theta}$, e.g., statistical test of Neyman-Pearson test, or minimax error probability. This is an immediate consequence of Proposition 2.2 and Lemma 2.3.

4 Covariant and contravariant channels

4.1 Universally optimal input states

Let $g \in \mathcal{G}$, where $\mathcal{G}$ is an element of compact Lie group or its discrete subgroup. Covariant and contravariant channels are those satisfying

$$\Lambda_\theta \circ \Upsilon_{U_g} = \Upsilon_{V_g} \circ \Lambda_\theta,$$

and

$$\Lambda_\theta \circ \Upsilon_{U_g} = \Upsilon_{\overline{V_g}} \circ \Lambda_\theta,$$

respectively. Here $g \to U_g$, $g \to V_g$ are representations of $\mathcal{G}$.

Example 4.1 *Let*
$$\Lambda_\theta^{\mathrm{cdep}} := \theta T + (1 - \theta)\,\Upsilon_m,$$

where $T(\rho) = \rho^T$, Υ_m is the channel which sends any input to the totally mixed state $\mathbf{1}/d$, and $\Theta := [0, 1/(d+1)] \subset \mathbb{R}$. Then $\Lambda_\theta^{\mathrm{cdep}}$ is completely positive, trace preserving, and contravariant.

Example 4.2 *Let Υ_c be the m to n optimal pure state cloner [9], which is covariant with $U_g := g^{\otimes m}$, $V_g := g^{\otimes n}$, $\mathcal{H}_{in} := \left(\mathbb{C}^d\right)^{\otimes_s m}$, and $\mathcal{H}_{out} := \left(\mathbb{C}^d\right)^{\otimes_s n}$. (Here, $\otimes_s$ denotes symmetric tensor product.)Then, the channels*

$$\Lambda_\theta^{\mathrm{cl}} := \theta \Upsilon_c + (1 - \theta)\,\Upsilon_m,\ \theta \in \Theta := [0, 1],$$

are covariant.

Example 4.3 *Another example is $\Lambda_{d,\theta}^{\mathrm{gp}}$ with $\mathcal{H}_{in} := \mathbb{C}^d$, and $\mathcal{H}_{out} := \mathbb{C}^d$,*

$$\Lambda_{d,\theta}^{\mathrm{gp}} := \sum_{j,k=0}^{d-1} \theta^{(j,k)} \Upsilon_{X_d^j Z_d^k},$$

where

$$\Theta := \left\{ \theta \,;\, \theta^{(j,k)} \geq 0, \sum_{j,k=0}^{d-1} \theta^{(j,k)} = 1 \right\},$$

and X_d, Z_d are generalized Pauli matrices defined by

$$X_d := \sum_{i=1}^{d-1} |i\rangle \langle i+1| + |d\rangle \langle 1|, \; Z_d := \sum_{i=1}^{d} e^{\frac{\sqrt{-1}2\pi i}{d}} |i\rangle \langle i|. \tag{5}$$

X_d and Z_d satisfy

$$(X_d)^d = (Z_d)^d = \mathbf{1}, \; e^{\frac{\sqrt{-1}2\pi i}{d}} Z_d X_d = X_d Z_d. \tag{6}$$

$\Lambda_{d,\theta}^{\mathrm{gp}}$ *is covariant with respect to*

$$\mathcal{G} = \mathcal{G}_d := \left\{ e^{\frac{\sqrt{-1}\,2\pi\,i}{d}} \, (X_d)^j \, (Z_d)^k \,;\, i,j,k = 0,1,\cdots,d-1 \right\}$$

and $U_g = V_g = g$. Indeed, if $\mathcal{H}_{in} = \mathcal{H}_{out} = \mathbb{C}^d$ and $U_g = V_g = g \in \mathcal{G}_d$, being covariant is equivalent to be a member of $\left\{ \Lambda_{d,\theta}^{\mathrm{gp}} \right\}$ [10].

Example 4.4 *An alternative parameterization of $\Lambda_{2,\theta}^{\mathrm{gp}}$ is given by*

$$\Lambda_{2,\eta}^{\mathrm{gp}} := \sum_{i=1}^{4} \Upsilon_{E_i},$$

where

$$E_1 := \begin{bmatrix} \eta^1 & 0 \\ 0 & \eta^2 \end{bmatrix}, \; E_2 := \begin{bmatrix} \eta^2 & 0 \\ 0 & \eta^1 \end{bmatrix},$$

$$E_3 := \begin{bmatrix} 0 & \eta^3 \\ \sqrt{1 - \sum_{i=1}^{3} (\eta^i)^2} & 0 \end{bmatrix},$$

$$E_4 := E_3^\dagger.$$

Example 4.5 *With $\mathcal{H}_{in} = \mathcal{H}_{out} = \mathbb{C}^2$, [6] had defined generalized damping channels :*

$$\Lambda_{p,\xi}^{\mathrm{damp}} = \sum_{i=1}^{4} \Upsilon_{F_i},$$

$$F_1 := \sqrt{p} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\xi} \end{bmatrix}, \; F_2 := \sqrt{1-p} \begin{bmatrix} \sqrt{\xi} & 0 \\ 0 & 1 \end{bmatrix},$$

$$F_3 := \sqrt{p} \begin{bmatrix} 0 & \sqrt{1-\xi} \\ 0 & 0 \end{bmatrix}, \; F_4 := \sqrt{1-p} \begin{bmatrix} 0 & 0 \\ \sqrt{1-\xi} & 0 \end{bmatrix}.$$

Then, $\left\{ \Lambda_{1/2,\xi}^{\mathrm{damp}} \right\}$ is covariant with respect to $U_g = V_g = g \in \mathcal{G}_2$. Indeed, $\left\{ \Lambda_{1/2,\xi}^{\mathrm{damp}} \right\}$ is a subset of $\left\{ \Lambda_{2,\theta}^{\mathrm{gp}} \right\}$.

Example 4.6 *Let*

$$\Lambda_{d,\theta}^{\mathrm{diag}} := \sum_{i=1}^{d} \Upsilon_{E_i},$$

where

$$E_1 := \mathrm{diag}\left(\theta^1, \theta^2, \cdots, \theta^{\lceil (d-1)/2 \rceil}, \sqrt{\sum_{i=1}^{\lceil (d-1)/2 \rceil} (\theta^i)^2}, 0, \cdots, 0\right),$$

$$E_i := X_d^{i-1} E_1 X_d^{i-1} \ (2 \le i \le d).$$

Then $\Lambda_{d,\xi}^{\mathrm{diag}}$ is covariant with respect to $U_g = V_g = g \in \mathcal{G}_d$: in fact, it turns out the family $\left\{\Lambda_{d,\xi}^{\mathrm{diag}}\right\}$ is a subset of the family $\{\Lambda_\theta^{\mathrm{gp}}\}$.

[5] and [6] had shown that a maximal entangles state $|\Phi_d\rangle$ is optimal for the family $\left\{\Lambda_{2,\theta}^{\mathrm{gp}}\right\}$ and also for the family $\left\{\Lambda_{1/2,\xi}^{\mathrm{damp}}\right\}$ in the following sense. For any input state $\rho_{in} \in \mathcal{B}\left(\mathcal{H}_{in} \otimes \mathcal{H}_R\right)$ and a measurement M over $\mathcal{B}\left(\mathcal{H}_{out} \otimes \mathcal{H}_R\right)$ with

$$\mathrm{E}\left[M, \Lambda_\theta \otimes \mathbf{I}\left(\rho_{in}\right)\right] = \theta, \tag{7}$$

there is a measurement M' such that

$$\mathrm{E}\left[M', \Lambda_\theta \otimes \mathbf{I}\left(|\Phi_d\rangle \langle \Phi_d|\right)\right] = \theta$$

and

$$\mathrm{V}\left[M, \Lambda_\theta \otimes \mathbf{I}\left(\rho_{in}\right)\right] = \mathrm{V}\left[M', \Lambda_\theta \otimes \mathbf{I}\left(|\Phi_d\rangle \langle \Phi_d|\right)\right].$$

Also, [11] studies discrimination of a pair of channels in $\left\{\Lambda_{d,\theta}^{\mathrm{gp}}\right\}$, and shows that $|\Phi_d\rangle$ minimizes Bayesian error probability for any prior distributions. In case of qubits, they extended their result to minimax error probability [13].

The following theorem is a generalization of these results. Below, we suppose the representation $g \to U_g$ occurs the decomposition

$$\mathcal{H}_{in} = \bigoplus_\mu \mathcal{H}_{in}^{(\mu)}, \ U_g = \bigoplus_\mu U_g^{(\mu)},$$

where $U_g^{(\mu)}$ acts on $\mathcal{H}_{in}^{(\mu)}$ and $g \to U_g^{(\mu)}$ is irreducible. Also, define $d_\mu := \dim \mathcal{H}_{in}^{(\mu)}$.

Theorem 4.7 *Consider the covariant or contravariant channel family $\{\Lambda_\theta\}_{\theta \in \Theta}$. Then, with $\mathcal{H}_R^{(\mu)} \simeq \mathcal{H}_{in}^{(\mu)}$ and*

$$\mathcal{H}_R = \bigoplus_\mu \mathcal{H}_R^{(\mu)},$$

the followidng input state is universally optimal:

$$|\psi_{opt}\rangle := c \bigoplus_\mu |\Phi_{d_\mu}\rangle, \tag{8}$$

where $|\Phi_{d_\mu}\rangle \in \mathcal{H}_{in}^{(\mu)} \otimes \mathcal{H}_R^{(\mu)}$, and c is the normalizing constant.

Proof. We state the proof only for covariant case, since the argument is almost parallel for contravariant case. Below, we compose a completely positive trace preserving map Γ_ψ with

$$\Gamma_\psi \left(\Lambda_\theta \otimes \mathbf{I}\left(|\psi_{opt}\rangle \langle\psi_{opt}|\right)\right) = \Lambda_\theta \otimes \mathbf{I}\left(|\psi\rangle \langle\psi|\right),$$

for an arbitrary $|\psi\rangle \in \mathcal{H}'_{in} \otimes \mathcal{H}'_R$, and use Proposition 2.2. Here,

$$\Lambda_\theta \otimes \mathbf{I}\left(|\psi_{opt}\rangle \langle\psi_{opt}|\right) \in \mathcal{H}_{out} \otimes \mathcal{H}_R,$$
$$\Lambda_\theta \otimes \mathbf{I}\left(|\psi\rangle \langle\psi|\right) \in \mathcal{H}_{out} \otimes \mathcal{H}'_R,$$

where $\mathcal{H}'_R \simeq \mathcal{H}_R$.

Γ_ψ is composed as follows; Prepare $|\psi\rangle$ in $\mathcal{H}'_{in} \otimes \mathcal{H}'_R$, where $\mathcal{H}'_{in} \simeq \mathcal{H}_{in}$. Apply the measurement M (defined later) jointly to $\mathcal{H}_R$-part of $\Lambda_\theta \otimes \mathbf{I}\left(|\psi_{opt}\rangle \langle\psi_{opt}|\right)$ and $\mathcal{H}'_{in}$-part of $|\psi\rangle$. Depending on the outcome $g \in \mathcal{G}$ of M, apply $V_g^\dagger$ to $\mathcal{H}_{out}$.

To define the measurement M, we first define the the state vector in $\mathcal{H}_R \otimes \mathcal{H}'_{in}$,

$$|\varphi_g\rangle := c' \bigoplus_\mu d_\mu \overline{U_g^{(\mu)}} \otimes \mathbf{1}_{\mathcal{H}_{in}^{(\mu)'}} |\Phi_{d_\mu}\rangle,$$

with the normalizing constant c', $U_g^{(\mu)}$ being in $\mathcal{H}_R^{(\mu)}$ and $\mathcal{H}_{in}^{(\mu)'} \simeq \mathcal{H}_{in}^{(\mu)}$. Then, the measurement M is the one which occurs state change

$$\rho \to c'' \mathbf{I}_{\mathcal{H}_{out} \otimes \mathcal{H}'_R} \otimes \Upsilon_{\langle\varphi_g|}\left(\rho\right),$$

with the probability density $\operatorname{tr} \mathbf{I}_{\mathcal{H}_{out} \otimes \mathcal{H}'_R} \otimes \Upsilon_{\langle\varphi_g|}\left(\rho\right)$. Here c'' is the normalizing constant, and the density is considered with respect to the Haar measure $\mathrm{d}g$ such that $\int_{\mathcal{G}} \mathrm{d}g = 1$.

In the end, we confirm that Γ_ψ meets the requirement. By composition, Γ_ψ is completely positive and trace preserving. Let $\{A_\kappa\}$ be the Kraus operators of Λ_θ. Also, let $\mathcal{H}'_R := \bigoplus_\mu \mathcal{H}_R^{(\mu)'}$ and $\mathcal{H}_R^{(\mu)'} \simeq \mathcal{H}_R^{(\mu)}$. Then, after the application of M and obtaining measurement result $g \in \mathcal{G}$, the state will be the mixture of

the pure state in $\mathcal{H}_{out} \otimes \mathcal{H}'_R$, such that

$$
\begin{aligned}
&\sqrt{c''} \left(\mathbf{1}_{\mathcal{H}_{out}} \otimes \langle \varphi_g | \otimes \mathbf{1}_{\mathcal{H}'_R} \right) \left(A_\kappa \otimes \mathbf{1}_{\mathcal{H}_R} \otimes \mathbf{1}_{\mathcal{H}'_{in}} \otimes \mathbf{1}_{\mathcal{H}'_R} | \psi_{opt} \rangle | \psi \rangle \right) \\
&= \sqrt{c''} \left(A_\kappa \otimes \mathbf{1}_{\mathcal{H}'_R} \right) \left(\mathbf{1}_{\mathcal{H}_{in}} \otimes \langle \varphi_g | \otimes \mathbf{1}_{\mathcal{H}'_R} \right) | \psi_{opt} \rangle | \psi \rangle \\
&= cc' \sqrt{c''} \left(A_\kappa \otimes \mathbf{1}_{\mathcal{H}'_R} \right) \bigoplus_\mu d_\mu \left(\mathbf{1}_{\mathcal{H}_{in}} \otimes \left(\langle \Phi_{d_\mu} | U_g^{(\mu)T} \otimes \mathbf{1}_{\mathcal{H}'_{in}} \right) \otimes \mathbf{1}_{\mathcal{H}_R^{(\mu)'}} \right) | \Phi_{d_\mu} \rangle | \psi \rangle \\
&= cc' \sqrt{c''} \left(A_\kappa \otimes \mathbf{1}_{\mathcal{H}'_R} \right) \bigoplus_\mu d_\mu \left(\mathbf{1}_{\mathcal{H}_{in}} \otimes \left(\langle \Phi_{d_\mu} | \mathbf{1}_{\mathcal{H}_R^{(\mu)}} \otimes U_g^{(\mu)} \right) \otimes \mathbf{1}_{\mathcal{H}_R^{(\mu)'}} \right) | \Phi_{d_\mu} \rangle | \psi \rangle \\
&= cc' \sqrt{c''} A_\kappa \bigoplus_\mu \sum_{i=1}^{d_\mu} | i \rangle_{\mathcal{H}_{in}^{(\mu)}} \otimes \left(\left({}_{\mathcal{H}_{in}^{(\mu)'}} \langle i | U_g^{(\mu)} \right) \otimes \mathbf{1}_{\mathcal{H}_R^{(\mu)'}} | \psi \rangle \right) \\
&= cc' \sqrt{c''} \left(A_\kappa U_g \right) \otimes \mathbf{1}_{\mathcal{H}'_R} | \psi \rangle .
\end{aligned}
$$

This mixture equals

$$
\begin{aligned}
&(cc')^2 \, c'' \left(\Lambda_\theta \circ \Upsilon_{U_g} \right) \otimes \mathbf{I}_{\mathcal{H}'_R} \left(| \psi \rangle \langle \psi | \right) \\
&= (cc')^2 \, c'' \left(\Upsilon_{V_g} \circ \Lambda_\theta \right) \otimes \mathbf{I}_{\mathcal{H}'_R} \left(| \psi \rangle \langle \psi | \right) .
\end{aligned}
$$

Therefore, applying $V_g^\dagger$ to $\mathcal{H}_{out}$, we have $\Lambda_\theta \otimes \mathbf{I}_{\mathcal{H}'_R} \left(| \psi \rangle \langle \psi | \right)$, as desired. ■

4.2 On $\Lambda_{1,\xi}^{\mathrm{damp}}$

[6] had shown that for $\Lambda_{1,\xi}^{\mathrm{damp}}$, $\rho = |2\rangle \langle 2| \in \mathcal{B}(\mathcal{H}_{in})$ is optimal for mean square error under the constraint (7). Despite this fact, $|2\rangle \langle 2|$ is *not* universally optimal as is shown below. Indeed,

$$
\begin{aligned}
&\left\| \Lambda_{1,\xi}^{\mathrm{damp}} (|2\rangle \langle 2|) - s \, \Lambda_{1,0}^{\mathrm{damp}} (|2\rangle \langle 2|) \right\|_1 = |1 - \xi - s| + \xi, \\
&\left\| \Lambda_{1,\xi}^{\mathrm{damp}} \otimes \mathbf{I} (|\Phi_2\rangle \langle \Phi_2|) - s \, \Lambda_{1,0}^{\mathrm{damp}} \otimes \mathbf{I} (|\Phi_2\rangle \langle \Phi_2|) \right\|_1 \\
&= \frac{1}{2} \sqrt{(1 - s + \xi)^2 + 4s\xi} + \frac{1}{2} |1 - \xi - s| .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{1}{2} &= \left\| \Lambda_{1,1/2}^{\mathrm{damp}} (|2\rangle \langle 2|) - \frac{1}{2} \Lambda_{1,0}^{\mathrm{damp}} (|2\rangle \langle 2|) \right\|_1 \\
&< \left\| \Lambda_{1,1/2}^{\mathrm{damp}} \otimes \mathbf{I} (|\Phi_2\rangle \langle \Phi_2|) - \frac{1}{2} \Lambda_{1,0}^{\mathrm{damp}} \otimes \mathbf{I} (|\Phi_2\rangle \langle \Phi_2|) \right\|_1 = \frac{\sqrt{2}}{2} .
\end{aligned}
$$

Therefore, by Lemma 2.4, we have the assertion.

11

4.3 An alternative proof for $\left\{\Lambda_{d,\theta}^{\mathrm{gp}}\right\}_{\theta\in\Theta}$

Given $\Lambda_{d,\theta}^{\mathrm{gp}}\otimes\mathbf{I}\left(|\Phi_d\rangle\langle\Phi_d|\right)$ and $\rho_{in}\in\mathcal{B}\left(\mathcal{H}\right)$, one can generate $\Lambda_{d,\theta}^{\mathrm{gp}}\otimes\mathbf{I}\left(\rho_{in}\right)$ in the following manner. Measure $\Lambda_\theta\otimes\mathbf{I}\left(|\Phi_d\rangle\langle\Phi_d|\right)$ by the projectors onto $\left\{X_d^j Z_d^k\otimes\mathbf{1}\,|\Phi_d\rangle\right\}_{j,k=0}^{d-1}$, and apply the unitary $X_d^j Z_d^k\otimes\mathbf{1}$ if (j,k) is observed.

This composition works also for any channel family $\{\Lambda_\theta^{\mathrm{ou}}\}$ with

$$\Lambda_\theta^{\mathrm{ou}}:=\frac{1}{d}\left(1-\sum_{i=1}^{d^2-1}\theta^i\right)\Upsilon_{U_1}+\sum_{i=2}^{d^2}\theta^i\Upsilon_{U_i},$$

$$\operatorname{tr}U_iU_j^\dagger=d\delta_{ij}.$$

5 Unital qubit channels

In this section, $\mathcal{H}_{in}=\mathcal{H}_{out}=\mathbb{C}^2$. Also we denote

$$Y_2:=\sqrt{-1}Z_2X_2.$$

and define

$$V=\left[\begin{array}{cc} e^{\sqrt{-1}b}\cos a & -e^{-\sqrt{-1}c}\sin a \\ e^{\sqrt{-1}c}\sin a & e^{-\sqrt{-1}b}\cos a \end{array}\right]\in\mathrm{SU}\left(2\right).$$

With $\boldsymbol{p}=(p_1,p_2)\ (p_1+p_2=1)$ and $V\in\mathrm{SU}\left(2\right)$, let

$$|\varphi_{\boldsymbol{p},V}\rangle:=\sqrt{p_1}\left(V\,|1\rangle\right)\otimes|1\rangle+\sqrt{p_2}\left(V\,|2\rangle\right)\otimes|2\rangle.$$

Also, let Γ_{UNOT} denote the universal not operation

$$\Gamma_{\mathrm{unot}}\left(C\right)=\overline{\Lambda_{Y_2}\left(C\right)},$$

which is positive trace preserving but not completely positive.

Observe that Y_2 is unitary and Hermite, and that

$$Y_2\,V\ =\overline{V}\,Y_2 \tag{9}$$

or equivalently,

$$\Upsilon_{Y_2}\circ\Gamma_{\mathrm{unot}}=\Gamma_{\mathrm{unot}}\circ\Upsilon_{Y_2}. \tag{10}$$

Lemma 5.1 *Suppose*

$$Y_2\,\Lambda_\theta\left(C\right)\,Y_2=\overline{\Lambda}_\theta\left(Y_2\,C\,Y_2\right), \tag{11}$$

or equivalently

$$\Lambda_\theta\circ\Gamma_{\mathrm{unot}}=\Gamma_{\mathrm{unot}}\circ\Lambda_\theta \tag{12}$$

Then, the input $|\Phi_2\rangle$ is universally optimal.

Proof. To use Proposition 2.2, we compose a trace preserving positive map Γ with

$$\Gamma\left(\Lambda_\theta \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right)\right) = \Lambda_\theta \otimes \mathbf{I}\left(|\varphi_{\boldsymbol{p},V}\rangle\langle\varphi_{\boldsymbol{p},V}|\right)$$

as follows. First, apply the unitary V^T to $\mathcal{H}_R$-part of $\Lambda_\theta \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right)$, obtaining

$$\Lambda_\theta \otimes \mathbf{I}\left(\left(\mathbf{1} \otimes V^T\right)|\Phi_2\rangle\langle\Phi_2|\left(\mathbf{1} \otimes V^{T\dagger}\right)\right)$$
$$= \Lambda_\theta \otimes \mathbf{I}\left(\left(V \otimes \mathbf{1}\right)|\Phi_2\rangle\langle\Phi_2|\left(V^\dagger \otimes \mathbf{1}\right)\right).$$

Second, measure $\mathcal{H}_R$-part by the measurement specified by the instrument

$$\left\{\sqrt{M}, \sqrt{\mathbf{1} - M}\right\},$$

where

$$M := p_1 |1\rangle\langle1| + p_2 |2\rangle\langle2|.$$

If the measurement result is the one corresponding to $\sqrt{M}$, then we are done. Otherwise, letting $\boldsymbol{p}' := (p_2, p_1)$, we obtain

$$\Lambda_\theta \otimes \mathbf{I}\left(|\varphi_{\boldsymbol{p}',V}\rangle\langle\varphi_{\boldsymbol{p}',V}|\right)$$
$$= \left(\Lambda_\theta \otimes \mathbf{I}\right) \circ \left(\Gamma_{\mathrm{unot}} \otimes \Gamma_{\mathrm{unot}}\right)\left(|\varphi_{\boldsymbol{p},V}\rangle\langle\varphi_{\boldsymbol{p},V}|\right)$$
$$= \left(\Gamma_{\mathrm{unot}} \otimes \Gamma_{\mathrm{unot}}\right) \circ \left(\Lambda_\theta \otimes \mathbf{I}\right)\left(|\varphi_{\boldsymbol{p},V}\rangle\langle\varphi_{\boldsymbol{p},V}|\right).$$

So we apply $\Gamma_{\mathrm{unot}} \otimes \Gamma_{\mathrm{unot}}$, to obtain $\Lambda_\theta \otimes \mathbf{I}\left(|\varphi_{\boldsymbol{p},V}\rangle\langle\varphi_{\boldsymbol{p},V}|\right)$. ∎

Any 4×4 Hermite matrix belongs to

$$\mathrm{span}_{\mathbb{R}}\left\{A \otimes B\,;\; A, B = \mathbf{1}, X_2, Y_2, Z_2\right\}.$$

So is Choi-Jamilokovski's representation $Ch\left(\Lambda_\theta\right)$. Since Λ_θ is trace preserving,

$$\mathrm{tr}_{\mathcal{H}_{out}} Ch\left(\Lambda_\theta\right) = \mathbf{1}_{in}.$$

Therefore, $Ch\left(\Lambda_\theta\right)$ is a positive element of $\mathrm{span}_{\mathbb{R}} \mathcal{TP}$, where

$$\mathcal{TP} := \left\{A \otimes B\,;\, A, B = \mathbf{1}, X_2, Y_2, Z_2, \text{ if } A = \mathbf{1}, \text{ then } B = \mathbf{1}\right\}$$

Lemma 5.2 *(12) holds if $Ch\left(\Lambda_\theta\right)$ is an element of*

$$\mathrm{span}_{\mathbb{R}}\left(\mathcal{TP} - \left\{X_2 \otimes \mathbf{1}, Y_2 \otimes \mathbf{1}, Z_2 \otimes \mathbf{1}\right\}\right),$$

or equivalently,

$$\Lambda_\theta\left(\mathbf{1}\right) = \mathbf{1}.$$

Proof. Since

$$\Lambda_\theta\left(C\right) = \mathrm{tr}_{\mathcal{H}_{in}} Ch\left(\Lambda_\theta\right)\left(\mathbf{1}_{out} \otimes C^T\right),$$

13

and

$$\overline{\Lambda_\theta}\left(ACA^\dagger\right) = \mathrm{tr}_{\mathcal{H}_{in}}\left(\mathbf{1}\otimes A^T\overline{Ch\left(\Lambda_\theta\right)}\mathbf{1}\otimes\overline{A}\right)\left(\mathbf{1}\otimes C^T\right),$$

$$A\Lambda_\theta\left(C\right)A^\dagger = \mathrm{tr}_{\mathcal{H}_{in}}\left(A\otimes\mathbf{1}\,Ch\left(\Lambda_\theta\right)\,A^\dagger\otimes\mathbf{1}\right)\left(\mathbf{1}\otimes C^T\right),$$

(11) is equivalent to

$$\mathbf{1}\otimes Y_2\,\overline{Ch\left(\Lambda_\theta\right)}\,\mathbf{1}\otimes Y_2 = Y_2\otimes\mathbf{1}\,Ch\left(\Lambda_\theta\right)Y_2\otimes\mathbf{1},$$

or equivalently, with $W = Ch\left(\Lambda_\theta\right)$,

$$W = \Gamma_{\mathrm{unot}}\otimes\Gamma_{\mathrm{unot}}\left(W\right). \tag{13}$$

Each element of $\mathcal{TP}$ other than $X_2\otimes\mathbf{1}$, $Y_2\otimes\mathbf{1}$ and $Z_2\otimes\mathbf{1}$ satisfies (13). Therefore, we have the assertion. ∎

Combining these lemmas, we have the following theorem.

Theorem 5.3 *Suppose $\mathcal{H}_{in} = \mathcal{H}_{out} = \mathbb{C}^2$. Then, the input $|\Phi_2\rangle$ is universally optimal if Λ_θ is unital.*

Example 5.4 *Due to (10), the family $\{\Upsilon_U\,;\,U\in\mathrm{SU}\left(2\right)\}$ satisfies (12).*

Example 5.5 *Channel family $\{\Lambda_\theta\}$ with*

$$Ch\left(\Lambda_\theta\right) = \begin{bmatrix} 1 & 0 & 0 & \theta^1 - \sqrt{-1}\theta^2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \theta^1 + \sqrt{-1}\theta^2 & 0 & 0 & 1 \end{bmatrix}$$

satisfies (12). In Kraus representation, Λ_θ is expressed as

$$\Lambda_\theta = \sum_{i=1}^{2}\Upsilon_{E_i},$$

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & \theta^1 + \sqrt{-1}\theta^2 \end{bmatrix}, E_2 = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{1 - \sum_{i=1}^{2}\left(\theta^i\right)^2} \end{bmatrix}.$$

6　A pair of measurements

Let us consider a family $\{\Lambda_\theta\}_{\theta\in\{+,-\}}$ such that $\Lambda_\theta : \mathcal{B}\left(\mathcal{H}_{in}\right)\to\mathcal{B}\left(\mathcal{H}_{out}\right)$, $\mathcal{H}_{out} = \mathbb{C}^m$, and

$$\Lambda_\theta\left(\rho\right) = \sum_{i=1}^{m}\left\{\mathrm{tr}\,\rho M_\theta\left(i\right)\right\}|i\rangle\langle i|. \tag{14}$$

This corresponds to measurements which outputs classical data "i" with probability $\mathrm{tr}\,\rho M_\theta\left(i\right)$.

Example 6.1 *Suppose*

$$M_+ (i) M_- (i) = 0, \ (i = 1, \cdots, m). \tag{15}$$

For example, suppose

$$\mathrm{rank} M_+ (i) = 1,$$
$$M_- (i) := \frac{1}{d-1} \{ \mathrm{tr} M_+ (i) \cdot \mathbf{1} - M_+ (i) \}. \tag{16}$$

Then,

$$M_- (i) M_+ (i) = M_+ (i) M_- (i)$$
$$= \frac{1}{d-1} \left\{ \mathrm{tr} M_+ (i) \cdot M_- (i) - (M_- (i))^2 \right\}$$
$$= 0$$

and

$$\sum_{i=1}^m M_- (i) = \frac{1}{d-1} \left\{ \mathrm{tr} \sum_{i=1}^m M_+ (i) \cdot \mathbf{1} - \sum_{i=1}^m M_+ (i) \right\}$$
$$= \frac{1}{d-1} \{ \mathrm{tr}\, \mathbf{1} \cdot \mathbf{1} - \mathbf{1} \} = \mathbf{1}.$$

Thus, (16) is a special case of (15).

An input $|\psi\rangle$ is universally optimal if

$$\sqrt{\rho_\psi} M_+ (i)^T \rho_\psi M_- (i)^T \sqrt{\rho_\psi} = 0, \ (i = 1, \cdots, m), \tag{17}$$

where ρ_ψ is as of (1). In particular, $|\Phi_d\rangle$ is universally optimal.

The proof is as follows. Suppose (17) holds. Then

$$\Lambda_+ \otimes \mathbf{I} (|\psi\rangle \langle\psi|) \Lambda_- \otimes \mathbf{I} (|\psi\rangle \langle\psi|)$$
$$= \sum_{i=1}^m |i\rangle \langle i| \otimes \sqrt{\rho_\psi} M_+ (i)^T \rho_\psi M_- (i)^T \sqrt{\rho_\psi}$$
$$= 0$$

Therefore, $\Lambda_+ \otimes \mathbf{I} (|\psi\rangle \langle\psi|)$ and $\Lambda_- \otimes \mathbf{I} (|\psi\rangle \langle\psi|)$ can be discriminated perfectly. Therefore, for any $\rho_{in} \in \mathcal{B} (\mathcal{H}_{in} \otimes \mathcal{H}_R)$, there is a trace preserving CPTP map Γ with

$$\Gamma (\Lambda_\theta \otimes \mathbf{I} (|\psi\rangle \langle\psi|)) = \Lambda_\theta \otimes \mathbf{I} (\rho_{in}) \quad (\theta = +, -),$$

and by Proposition 2.2, we have the assertion.

Example 6.2 *Let $\{U_i\}_{i=1}^m$ be unitary matrices such that*

$$\sum_{i=1}^m U_i A U_i^\dagger = c \, (\mathrm{tr}\, A) \, \mathbf{1}, \tag{18}$$

15

and define

$$M_\theta(i) := \frac{1}{c} U_i M_\theta U_i^\dagger, \tag{19}$$

where

$$[M_+, M_-] = 0,$$
$$M_\theta \geq 0,$$
$$\operatorname{tr} M_+ = \operatorname{tr} M_- = 1.$$

Then, $|\Phi_d\rangle$ is universally optimal.
    The proof is as follows. Observe

$$[\Lambda_+ \otimes \mathbf{I}\,(|\Phi_d\rangle\,\langle\Phi_d|)\,, \Lambda_- \otimes \mathbf{I}\,(|\Phi_d\rangle\,\langle\Phi_d|)]$$
$$= \frac{1}{d^2 c^2} \sum_{i=1}^m |i\rangle\,\langle i| \otimes \left[\left(U_i M_+ U_i^\dagger \right)^T, \left(U_i M_- U_i^\dagger \right)^T \right]$$
$$= 0.$$

Also,

$$\| \Lambda_+ \otimes \mathbf{I}\,(|\varphi\rangle\,\langle\varphi|) - s\,\Lambda_- \otimes \mathbf{I}\,(|\varphi\rangle\,\langle\varphi|) \|_1$$
$$= \sum_{i=1}^m \left\| \sqrt{\rho_\varphi} \left(M_+(i)^T - s\,M_-(i)^T \right) \sqrt{\rho_\varphi} \right\|_1$$
$$\leq \sum_{i=1}^m \left\| \sqrt{\rho_\varphi} \left| M_+(i)^T - s\,M_-(i)^T \right| \sqrt{\rho_\varphi} \right\|_1$$
$$= \sum_{i=1}^m \operatorname{tr} \rho_\varphi \left| M_+(i) - s\,M_-(i) \right|^T$$
$$= \frac{1}{c} \operatorname{tr} \rho_\varphi \left(\sum_{i=1}^m U_i \left| M_+ - s\,M_- \right| U_i^\dagger \right)^T$$
$$= \operatorname{tr} \rho_\varphi \operatorname{tr} \left| M_+ - s\,M_- \right| = \operatorname{tr} \left| M_+ - s\,M_- \right|,$$

where the inequality in the third line is true if $|\varphi\rangle = |\Phi_d\rangle$. Therefore, by Lemma 2.4, we have the assertion.

Example 6.3 *Let*

$$M_\theta(2) = \mathbf{1} - M_\theta(1),$$
$$M_-(2) = M_+(1) = M = \operatorname{diag}(a_1, \cdots, a_d),$$
$$a_1 > a_2 \geq \cdots \geq a_d.$$

16

Then,

$$\|\Lambda_+ \otimes \mathbf{I}\left(|\psi\rangle\langle\psi|\right) - s\,\Lambda_- \otimes \mathbf{I}\left(|\psi\rangle\langle\psi|\right)\|_1$$
$$= \left\|\sqrt{\rho_\psi}\left(-s\mathbf{1} + (1+s)\,M^T\right)\sqrt{\rho_\psi}\right\|_1 + \left\|\sqrt{\rho_\psi}\left(\mathbf{1} - (1+s)\,M^T\right)\sqrt{\rho_\psi}\right\|_1$$
$$\leq \operatorname{tr}\rho_\psi\left(\left|-s\mathbf{1} + (1+s)\,M^T\right| + \left|\mathbf{1} - (1+s)\,M^T\right|\right)$$
$$= \sum_{i=1}^{d}\rho_{\psi,i,i}\left(\left|-s + (1+s)\,a_i\right| + \left|1 - (1+s)\,a_i\right|\right)$$
$$= \sum_{i=1}^{d}\rho_{\psi,i,i}\left(\left|(1-a_i)\,s - a_i\right| + \left|a_i s - (1-a_i)\right|\right)$$
$$\leq \left|(1-a_1)\,s - a_1\right| + \left|a_1 s - (1-a_1)\right|,$$

and the inequalities in the third and the fourth line are achieved by $\rho_\psi = |1\rangle\langle1|$. Therefore, by Lemma 2.4, $|\psi\rangle = |1\rangle\,|1\rangle$ is universally optimal.

7   SU(d) family

7.1 $d = 2$ case

In this subsection, we consider the family $\{\Upsilon_U\,;\,U \in \mathrm{SU}\,(2)\}$. [4] had shown that $|\Phi_2\rangle$ is optimal for the mean square error with the constraint (7). Also, as stated in Theorem 5.3, Section 5, $|\Phi_2\rangle$ is a universally optimal state.

Define for $|\psi\rangle \in \mathcal{H}_{in} \otimes \mathcal{H}_R$,

$$\mathcal{U}\left(\psi\right) := \{U;\operatorname{tr}\rho_\psi U = 0,\, U \in \mathrm{SU}\,(d)\}.$$

Below, we consider the test between the two hypotheses, $U = \mathbf{1}$ v.s. $U \in \mathcal{U}\,(\psi)$. In other words, $\mathcal{D} = \{0,1\}$ and the loss function l^ψ is such that

$$l_U^\psi(0) = \begin{cases} 1, & (U \in \mathcal{U}\,(\psi)) \\ 0, & (U = \mathbf{1}) \end{cases},$$
$$l_U^\psi(1) = \begin{cases} 0, & (U \in \mathcal{U}\,(\psi)) \\ 1, & (U = \mathbf{1}) \end{cases},$$
$$l_U^\psi(1) = l_U^\psi(0) = 0\,,\, U \notin \mathcal{U}\,(\psi) \cup \{\mathbf{1}\}$$

Theorem 7.1 *Consider the family $\{\Upsilon_U\,;\,U \in \mathrm{SU}\,(2)\}$. Then, $|\Phi_2\rangle$ is strictly universally optimal.*

Proof. Consider the loss function l^{Φ_2}. Then, since $|\Phi_2\rangle$ and $U \otimes \mathbf{1}\,|\Phi_2\rangle$ $(U \in \mathcal{U}\,(\Phi_2))$ are orthogonal, for any prior distribution π,

$$\min_M R\left(l^{\Phi_2}, M, \pi, |\Phi_2\rangle\right) = 0.$$

This is not the case if the input $|\psi\rangle$ is not maximally entangled. Indeed, any $U \in \mathrm{SU}\,(2)$ satisfy $|U_{11}| = |U_{22}|$. Without loss of generality, suppose the Schmidt

basis of $|\psi\rangle = \sqrt{p_1}\,|1\rangle\,|1\rangle + \sqrt{p_2}\,|2\rangle\,|2\rangle$, where $p_1 \neq p_2$. Then, the inner product between $|\psi\rangle$ and $U \otimes \mathbf{1}\,|\psi\rangle$ equals

$$\langle\psi|\,U \otimes \mathbf{1}\,|\psi\rangle = p_1 U_{11} + p_2 U_{22}.$$

But this cannot equal to 0 because of $p_1 \neq p_2$ and $|U_{11}| = |U_{22}|$. Therefore,

$$\min_M R\left(l^{\Phi_2}, M, \pi, |\psi\rangle\right) \neq 0.$$

∎

7.2   Tests on $\mathrm{SU}\,(d)$ $(d \geq 3)$

This subsection is devoted to the proof of the following theorem.

Theorem 7.2 *Consider the channel family $\{\Upsilon_U\,;\,U \in \mathrm{SU}\,(d)\}$, where $d \geq 3$. Then, any $|\psi\rangle \in \mathcal{H}_{in} \otimes \mathcal{H}_R$ is strictly admissible.*

First, we introduce a series of propositions and lemmas.

Proposition 7.3 *Consider the channel family $\{\Upsilon_U; U \in \mathrm{SU}\,(d)\}$. Then, if $\mathcal{U}\,(\psi) - \mathcal{U}\,(\psi') \neq \emptyset$,*

$$\min_M R\left(l^\psi, M, \pi, |\psi'\rangle\right) > \min_M R\left(l^\psi, M, \pi, |\psi\rangle\right), \exists\pi.$$

Proof. By the definition of $\mathcal{U}\,(\psi)$, there is a projective binary measurement $\{M_0, M_1\}$ which distinguishes $|\psi\rangle$ and $\{U \otimes \mathbf{1}\,|\psi\rangle\,;\,U \in \mathcal{U}\,(\psi)\}$ without error. Therefore, by the definition of l^ψ,

$$\min_M R\left(l^\psi, M, \pi, |\psi\rangle\right) = 0, \forall\pi.$$

If $U \in \mathcal{U}\,(\psi) - \mathcal{U}\,(\psi') \neq \emptyset$, then $|\psi'\rangle$ and $U \otimes \mathbf{1}\,|\psi'\rangle$ can not be distinguished perfectly. Therefore,

$$\min_M R\left(l^\psi, M, \pi, |\psi'\rangle\right) > 0, \exists\pi.$$

Therefore, we have the assertion. ∎

For $x \in \mathbb{R}^d$, define

$$Ang\,(x) := \left\{\vec{\omega}; \vec{\omega} \in \mathbb{C}^{d-1}, |\omega_i| = 1, \sum_{i=1}^{d-1} x_i\omega_i + x_d = 0\right\}.$$

Also, if

$$x_{i_1} \geq x_{i_2} \geq \cdots \geq x_{i_d},$$

$$x_j^\downarrow := x_{i_j},\ x^\downarrow := \left(x_1^\downarrow, x_2^\downarrow, \cdots, x_d^\downarrow\right).$$

The proof of the following lemma will be given in Appendix A

Lemma 7.4 *Suppose* $d \geq 3$, $x_i, x_i' > 0$ $(i = 1, \cdots, d)$, $x_1^{\downarrow} > \sum_{i=2}^{d} x_i^{\downarrow}$, *and* $x_1'^{\downarrow} > \sum_{i=2}^{d} x_i'^{\downarrow}$. *If* $Ang(x) \subset Ang(x')$, *then* $x' = sx$ *for some* $s \in \mathbb{R}$.

Lemma 7.5 *Let* $\rho_0, \rho_1 \in \mathcal{B}(\mathbb{C}^d)$, *and* $\{|i\rangle\}_{i=1}^{d}$ *be an orthonormal basis in* $\mathbb{C}^d$. *Suppose an Hermitian matrix* A *satisfies*

$$\langle i| U^{\dagger} A U |i\rangle = 0, (i = 0, \cdots, d),$$

for any unitary $U \in U(d)$ *such that*

$$0 < \langle i| U^{\dagger} \rho_j U |i\rangle < \frac{1}{2} \ (i = 1, \cdots, d, \ j = 0, 1). \tag{20}$$

Then we have

$$A = 0.$$

Proof. Let

$$E_{ij} := |i\rangle \langle j| + |j\rangle \langle i|, \ F_{ij} := \sqrt{-1}(|i\rangle \langle j| - |j\rangle \langle i|).$$

Then, there are real numbers a_i, b_{ij}, and c_{ij} with

$$U^{\dagger} A U = \sum_i a_i |i\rangle \langle i| + \sum_{i>j} (b_{ij} E_{ij} + c_{ij} F_{ij}).$$

Due to $\langle i| U^{\dagger} A U |i\rangle = 0$, $a_i = 0$.

If $U \in U(d)$ satisfies (20), then any member of neighborhood of U satisfies (20). Therefore,

$$\langle i| [U^{\dagger} A U, H] |i\rangle = 0, (i = 0, \cdots, d),$$

for any Hermitian matrix H. Also,

$$[E_{ij}, F_{ij}] = -2\sqrt{-1}(|i\rangle \langle i| + |j\rangle \langle j|),$$

and, if $\delta_{ik}\delta_{jl} = \delta_{il}\delta_{jk} = 0$,

$$\langle m| [E_{ij}, F_{kl}] |m\rangle = 0.$$

Therefore,

$$\langle i| [U^{\dagger} A U, E_{ij}] |i\rangle = -2\sqrt{-1}c_{ij},$$
$$\langle i| [U^{\dagger} A U, F_{ij}] |i\rangle = 2\sqrt{-1}b_{ij}.$$

Hence, $b_{ij} = c_{ij} = 0$. After all, $U^{\dagger} A U = 0$, implying $A = 0$. ∎

Now, we are in the position to present the proof of Theorem 7.2.

Proof. (Theorem 7.2) In view of Proposition 7.3, it suffices to show that $\mathcal{U}(\psi) \subset \mathcal{U}(\psi')$ implies $\rho_{\psi'} = \rho_{\psi}$.

Suppose $\mathcal{U}(\psi) \subset \mathcal{U}(\psi')$. Let us define, for $x \in \mathbb{R}^d$,

$$\widetilde{Ang}(x) := \left\{ \vec{\omega}; \vec{\omega} \in \mathbb{C}^d, |\omega_i| = 1, \sum_{i=1}^{d} \omega_i x_i = 0, \prod_{i=1}^{d} \omega_i = 1 \right\},$$

and, for a $d \times d$ matrix A,

$$\mathrm{Diag}(A) := (A_{11}, A_{22}, \cdots, A_{dd}).$$

Then,

$$\mathcal{U}(\rho_\psi) = \left\{ U \mathrm{diag}(\vec{\omega}) U^\dagger; U \in U(d), \vec{\omega} \in \widetilde{Ang}\left(\mathrm{Diag}\left(U^\dagger \rho_\psi U\right)\right) \right\}.$$

Therefore, $\mathcal{U}(\psi) \subset \mathcal{U}(\psi')$ implies

$$\widetilde{Ang}\left(\mathrm{Diag}\left(U^\dagger \rho_\psi U\right)\right) \subset \widetilde{Ang}\left(\mathrm{Diag}\left(U^\dagger \rho_{\psi'} U\right)\right),$$

or equivalently

$$Ang\left(\mathrm{Diag}\left(U^\dagger \rho_\psi U\right)\right) \subset Ang\left(\mathrm{Diag}\left(U^\dagger \rho_{\psi'} U\right)\right)$$

for all $U \in \mathrm{SU}(d)$.

Here we have recourse to Lemma 7.4, with $x_i := \langle i | U^\dagger \rho_\psi U | i \rangle$ and $x_i' := \langle i | U^\dagger \rho_{\psi'} U | i \rangle$. Suppose $x_1^\downarrow > \sum_{i=2}^{d} x_i^\downarrow$, and $x_1'^\downarrow > \sum_{i=2}^{d} x_i'^\downarrow$. Due to $\mathrm{tr}\, U^\dagger \rho_\psi U = \mathrm{tr}\, U^\dagger \rho_{\psi'} U = 1$, this is equivalent to

$$0 < \langle i | U^\dagger \rho_\psi U | i \rangle < \frac{1}{2},$$

$$0 < \langle i | U^\dagger \rho_{\psi'} U | i \rangle < \frac{1}{2}, \quad (i = 1, \cdots, d).$$

Therefore, by Lemma 7.4,

$$\mathrm{Diag}\left(U^\dagger \rho_\psi U\right) = \mathrm{Diag}\left(U^\dagger \rho_{\psi'} U\right),$$

which leads to

$$\mathrm{Diag}\left(U^\dagger \left(\rho_\psi - \rho_{\psi'}\right) U\right) = 0.$$

Therefore, by Lemma 7.5, $\rho_\psi = \rho_{\psi'}$. Thus we have the assertion. ∎

8 Universal enhancement by entanglement

8.1 Subfamily of $\left\{ \Lambda_{2,\theta}^{\mathrm{gp}} \right\}$

Observe, in Example 4.6, a separable state

$$|\psi\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle_{in} |f\rangle_R \succeq^c |\Phi_d\rangle,$$

where $|f\rangle_R \in \mathcal{H}_R$ is arbitrary, is universally optimal. Combining with the fact that $|\Phi_d\rangle$ is universally optimal (Theorem 4.7), we have

$$|\psi\rangle \equiv^c |\Phi_d\rangle. \tag{21}$$

The proof is as follows. For unitary operators $U_i \in U(\mathcal{H}_R)$ $(i = 1, \cdots, d)$, let $U_1 \oplus \cdots \oplus U_d$ be the unitary operator acting on $\mathcal{H}_{in} \otimes \mathcal{H}_R = \mathcal{H}_{out} \otimes \mathcal{H}_R$ such that

$$(U_1 \oplus \cdots \oplus U_d) |i\rangle_{in} |j\rangle_R = |i\rangle_{in} U_i |j\rangle_R.$$

Then, if $U_i |f\rangle = |i\rangle$,

$$(U_1 \oplus \cdots \oplus U_d) |\psi\rangle = |\Phi_d\rangle.$$

Observe

$$\left(\Lambda_{d,\theta}^{\mathrm{diag}} \otimes \mathbf{I}\right) \circ \Upsilon_{U_1 \oplus \cdots \oplus U_d} = \Upsilon_{U_1 \oplus \cdots \oplus U_d} \circ \left(\Lambda_{d,\theta}^{\mathrm{diag}} \otimes \mathbf{I}\right).$$

Therefore,

$$\Upsilon_{U_1 \oplus \cdots \oplus U_d} \circ \left(\Lambda_{d,\theta}^{\mathrm{diag}} \otimes \mathbf{I}\right)(|\psi\rangle \langle \psi|) = \left(\Lambda_{d,\theta}^{\mathrm{diag}} \otimes \mathbf{I}\right) \circ \Upsilon_{U_1 \oplus \cdots \oplus U_d}(|\psi\rangle \langle \psi|)$$
$$= \left(\Lambda_{d,\theta}^{\mathrm{diag}} \otimes \mathbf{I}\right)(|\Phi_d\rangle \langle \Phi_d|).$$

Therefore, by proposition 2.2, we have the assertion (21).

So in this case, entanglement between $\mathcal{H}_{in}$ and $\mathcal{H}_R$ is not necessary. When an entangled state is strictly universally better than any separable state? Below, the condition for $|\Phi_d\rangle$ to be strictly universally better than any separable states is studied. After investigating the subfamily of $\left\{\Lambda_{2,\theta}^{\mathrm{gp}}\right\}$ in this subsection, we move to families of measurements (Examples 6.1 and 6.1) in the next subsection.

Theorem 8.1 *With $\xi_\theta \in \mathbb{R}^3$, let $\Lambda_\theta = \Lambda_{2,\xi_\theta}^{\mathrm{gp}}$. Then, there is a separable state $|\psi\rangle = |\psi_{in}\rangle |\psi_R\rangle$ with $|\psi_{in}\rangle |\psi_R\rangle \succeq^c |\Phi_2\rangle$, or equivalently*

$$\{\Lambda_\theta(|\psi_{in}\rangle \langle \psi_{in}|)\}_{\theta \in \Theta} \succeq^c \{\Lambda_\theta \otimes \mathbf{I}(|\Phi_2\rangle \langle \Phi_2|)\}_{\theta \in \Theta} \tag{22}$$

if and only if

(i) $\{\xi_\theta\}_{\theta \in \Theta}$ *is on a straight line.*

(ii) *If there is at least a pair θ_1, θ_2 such that $\xi_+ := \xi_{\theta_1}$ and $\xi_- := \xi_{\theta_2}$ are distinct, ξ_+, ξ_- satisfies*

$$\xi_+^1 \xi_-^0 = \xi_-^1 \xi_+^0, \ \xi_+^2 \xi_-^3 = \xi_-^2 \xi_+^3, \tag{23}$$

or

$$\xi_+^2 \xi_-^0 = \xi_-^2 \xi_+^0, \ \xi_+^3 \xi_-^1 = \xi_-^3 \xi_+^1, \tag{24}$$

or

$$\xi_+^3 \xi_-^0 = \xi_-^3 \xi_+^0, \ \xi_+^1 \xi_-^2 = \xi_-^1 \xi_+^2, \tag{25}$$

where $\xi_\pm^0 := 1 - \xi_\pm^1 - \xi_\pm^2 - \xi_\pm^3$.

For example, $\left\{\Lambda_{\frac{1}{2},\xi}^{\mathrm{damp}}\right\}$ in Example 4.5 does not satisfy (i). Therefore, $|\Phi_2\rangle$ is strictly universally better than any separable states. On the other hand, $\left\{\Lambda_{2,\xi}^{\mathrm{diag}}\right\}$ (Example 4.6) satisfies the hypothesis of the theorem. Therefore, there is a separable state which is as good as $|\Phi_2\rangle$.

Proof. We first study the case where $\Theta = \{+,-\}$ and $\Lambda_+ = \Lambda_{2,\xi_+}^{\mathrm{gp}}$ and $\Lambda_- = \Lambda_{2,\xi_-}^{\mathrm{gp}}$. and give necessary and sufficient conditions for (22).

Suppose (22) holds. Then, since $\Lambda_+ \otimes \mathbf{I}\,(|\Phi_2\rangle\langle\Phi_2|)$ and $\Lambda_- \otimes \mathbf{I}\,(|\Phi_2\rangle\langle\Phi_2|)$ commutes, by Lemma 2.5, $\Lambda_+\,(|\psi_{in}\rangle\langle\psi_{in}|)$ and $\Lambda_-\,(|\psi_{in}\rangle\langle\psi_{in}|)$ has to commute. Let $\vec{r}$ and $\vec{r}_\theta$ be a Bloch vector of $|\psi_{in}\rangle\langle\psi_{in}|$ and $\Lambda_\theta\,(|\psi_{in}\rangle\langle\psi_{in}|)$, respectively. Then, this means that

$$\vec{r}_- = \alpha \vec{r}_+, \tag{26}$$

for a real number α. Also,

$$\vec{r}_\theta = \mathrm{diag}\left(a_\theta^1, a_\theta^2, a_\theta^3\right)\,\vec{r},$$

where

$$a_\theta^1 := 1 - 2\xi_\theta^2 - 2\xi_\theta^3,$$
$$a_\theta^2 := 1 - 2\xi_\theta^1 - 2\xi_\theta^3,$$
$$a_\theta^3 := 1 - 2\xi_\theta^1 - 2\xi_\theta^2.$$

Let us denote by $\rho\,(\vec{r})$ the state with Bloch vector $\vec{r}$. By simple calculations, we can verify

$$\|\rho\,(\vec{r}_+) - s\,\rho\,(\alpha \vec{r}_+)\|_1$$
$$= \left|\frac{1}{2} + \|\vec{r}_+\| - s\left(\frac{1}{2} + \alpha\,\|\vec{r}_+\|\right)\right| + \left|\frac{1}{2} - \|\vec{r}_+\| - s\left(\frac{1}{2} - \alpha\,\|\vec{r}_+\|\right)\right|$$

is non-decreasing in $\|\vec{r}_+\|$ for any $s \geq 0$. Therefore, by Lemma 2.4,

$$\{\rho\,(\vec{r}_+), \rho\,(\alpha\vec{r}_+)\} \succeq^c \left\{\rho\,(\vec{r}_+'), \rho\,(\alpha\vec{r}_+')\right\},$$

if and only if $\|\vec{r}_+\| \geq \|\vec{r}_+'\|$.

Therefore, we concentrate on $\vec{r}$ which maximizes $\|\vec{r}_+\| = \|\mathrm{diag}\left(a_+^1, a_+^2, a_+^3\right)\,\vec{r}\|$. This maximum can be achieved at $\vec{r} = (1,0,0)$, $(0,1,0)$, or $(0,0,1)$. Therefore, (22) holds if and only if

$$\left\{\rho\left(a_+^1,0,0\right), \rho\left(a_-^1,0,0\right)\right\} \succeq^c \left\{\Lambda_\theta \otimes \mathbf{I}\,(|\Phi_2\rangle\langle\Phi_2|)\right\}_{\theta\in\{+,-\}}, \tag{27}$$

or

$$\left\{\rho\left(0,a_+^2,0\right), \rho\left(0,a_-^2,0\right)\right\} \succeq^c \left\{\Lambda_\theta \otimes \mathbf{I}\,(|\Phi_2\rangle\langle\Phi_2|)\right\}_{\theta\in\{+,-\}}, \tag{28}$$

or

$$\left\{\rho\left(0,0,a_+^3\right), \rho\left(0,0,a_-^3\right)\right\} \succeq^c \left\{\Lambda_\theta \otimes \mathbf{I}\,(|\Phi_2\rangle\langle\Phi_2|)\right\}_{\theta\in\{+,-\}}. \tag{29}$$

Suppose (27) is the case. Then, in view of Lemma 2.4, we have to have

$$
\begin{aligned}
&\left\| \rho\left(a_+^1, 0, 0\right) - s\,\rho\left(a_-^1, 0, 0\right) \right\|_1 \\
&= \left| \xi_+^0 + \xi_+^1 - s\left(\xi_-^0 + \xi_-^1\right) \right| + \left| \xi_+^2 + \xi_+^3 - s\left(\xi_-^2 + \xi_-^3\right) \right| \\
&\geq \left| \xi_+^0 - s\xi_-^0 \right| + \left| \xi_+^1 - s\xi_-^1 \right| + \left| \xi_+^2 - s\xi_-^2 \right| + \left| \xi_+^3 - s\xi_-^3 \right| \\
&= \left\| \Lambda_+ \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right) - s\,\Lambda_- \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right) \right\|_1 ,
\end{aligned}
\tag{30}
$$

where

$$
\xi_\pm^0 := 1 - \xi_\pm^1 - \xi_\pm^2 - \xi_\pm^3 .
$$

On the other hand, observe

$$
\begin{aligned}
\left| \xi_+^0 + \xi_+^1 - s\left(\xi_-^0 + \xi_-^1\right) \right| &\leq \left| \xi_+^0 - s\xi_-^0 \right| + \left| \xi_+^1 - s\xi_-^1 \right| , \\
\left| \xi_+^2 + \xi_+^3 - s\left(\xi_-^2 + \xi_-^3\right) \right| &\leq \left| \xi_+^2 - s\xi_-^2 \right| + \left| \xi_+^3 - s\xi_-^3 \right| .
\end{aligned}
$$

Therefore, the inequality (2) is true for any $s \geq 0$ if and only if identities in above two inequalities hold for any $s \geq 0$. Therefore, (27) if and only if (23).

Similarly, (28) and (29) holds if and only if (24) and (25), respectively. Therefore, in the case of $\Theta = \{+, -\}$, there is $|\psi_{in}\rangle$ with (22) if and only if one of (23), (24) or (25) holds.

Next, we treat the case where Θ is an arbitrary set, and $\Lambda_\theta = \Lambda_{2,\xi_\theta}^{\mathrm{gp}}$. We suppose that there is at least a pair θ_1, θ_2 such that $\xi_+ := \xi_{\theta_1}$ and $\xi_- := \xi_{\theta_2}$ are distinct. In view of Lemma 2.5, (22) holds only if $\Lambda_\theta\left(|\psi_{in}\rangle\langle\psi_{in}|\right)$ and $\Lambda_{\theta'}\left(|\psi_{in}\rangle\langle\psi_{in}|\right)$ commutes for any θ, θ'. Therefore, $\{\Lambda_\theta\left(|\psi_{in}\rangle\langle\psi_{in}|\right)\}_{\theta\in\Theta}$ is on a straight line passing through origin. Denoting Λ_{θ_1} and Λ_{θ_2} by Λ_+ and Λ_-, respectively, for any $\theta \in \Theta$, there is $\lambda_\theta \in \mathbb{R}$ such that

$$
\Lambda_\theta\left(|\psi_{in}\rangle\langle\psi_{in}|\right) = \lambda_\theta \Lambda_+\left(|\psi_{in}\rangle\langle\psi_{in}|\right) + \left(1 - \lambda_\theta\right)\Lambda_-\left(|\psi_{in}\rangle\langle\psi_{in}|\right) .
\tag{31}
$$

We assert (22) holds if and only if

$$
\begin{aligned}
&\left\{ \Lambda_+\left(|\psi_{in}\rangle\langle\psi_{in}|\right), \Lambda_-\left(|\psi_{in}\rangle\langle\psi_{in}|\right) \right\}_{\theta\in\Theta} \\
&\succeq^c \left\{ \Lambda_+ \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right), \Lambda_- \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right) \right\}
\end{aligned}
\tag{32}
$$

and

$$
\Lambda_\theta = \lambda_\theta \Lambda_+ + \left(1 - \lambda_\theta\right)\Lambda_- .
\tag{33}
$$

The statement of the present theorem follows immediately from this assertion.

First, we show 'only if '. Obviously, (22) implies (32). Also, due to (22), for any positive operator $F \leq \mathbf{1}$ there is a positive operator $F' \leq \mathbf{1}$ such that

$$
\begin{aligned}
&\mathrm{tr}\left[\left\{\lambda_\theta\Lambda_+ + \left(1 - \lambda_\theta\right)\Lambda_- - \Lambda_\theta\right\} \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right)\right] F \\
&= \mathrm{tr}\left[\left\{\lambda_\theta\Lambda_+ + \left(1 - \lambda_\theta\right)\Lambda_- - \Lambda_\theta\right\}\left(|\psi_{in}\rangle\langle\psi_{in}|\right)\right] F' \\
&= 0.
\end{aligned}
$$

Here the second identity is due to (31). Since $F \leq \mathbf{1}$ is arbitrary, we have

$$
\Lambda_\theta \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right) = \left\{\lambda_\theta\Lambda_+ + \left(1 - \lambda_\theta\right)\Lambda_-\right\} \otimes \mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right) .
$$

23

Therefore, $\{\Lambda_\theta \otimes \mathbf{I}(|\Phi_0\rangle\langle\Phi_0|)\}_{\theta\in\Theta}$ is also on a straight line, and so is $\{\Lambda_\theta\}_{\theta\in\Theta}$. Thus we have (33).

To show the opposite, suppose (32) and (33) holds. Then, for any measurement M, there exists a measurement M' such that

$$
\begin{aligned}
P^M_{\Lambda_\theta\otimes\mathbf{I}(|\Phi_2\rangle\langle\Phi_2|)} &= \lambda_\theta\, P^M_{\Lambda_+\otimes\mathbf{I}(|\Phi_2\rangle\langle\Phi_2|)} + (1-\lambda_\theta)\, P^M_{\Lambda_-\otimes\mathbf{I}(|\Phi_2\rangle\langle\Phi_2|)} \\
&= \lambda_\theta\, P^{M'}_{\Lambda_+(|\psi_{in}\rangle\langle\psi_{in}|)} + (1-\lambda_\theta)\, P^{M'}_{\Lambda_-(|\psi_{in}\rangle\langle\psi_{in}|)} \\
&= P^{M'}_{\Lambda_\theta(|\psi_{in}\rangle\langle\psi_{in}|)}.
\end{aligned}
$$

Hence, we have (22), and our assertion is proved. Thus, we have Theorem 8.1. ∎

8.2 A pair of measurements

In this subsection, we investigate the measurement families studied in Examples 6.1, 6.2 of Section 6.

First, in Example 6.1, $|\psi_{in}\rangle|\psi_R\rangle \succeq^c |\Phi_d\rangle$ holds if and only if its output can be discriminated with certainty, or equivalently,

$$
\langle\psi_{in}|\,M_+(i)\,|\psi_{in}\rangle\,\langle\psi_{in}|\,M_-(i)\,|\psi_{in}\rangle = 0, \quad i = 1,\cdots,m. \tag{34}
$$

Proposition 8.2 *In case of (16), (34) holds if and only if either*

$$
M_+(i)\,|\psi_{in}\rangle = 0 \tag{35}
$$

or

$$
M_+(i) = c\,|\psi_{in}\rangle\langle\psi_{in}| \quad (c: constant) \tag{36}
$$

holds for any i.

Proof. If $\langle\psi_{in}|\,M_+(j)\,|\psi_{in}\rangle = 0$, we have (35). On the other hand, suppose $\langle\psi_{in}|\,M_+(i)\,|\psi_{in}\rangle \neq 0$. Then, for (34) to be true, $\langle\psi_{in}|\,M_-(i)\,|\psi_{in}\rangle = 0$ has to hold. Therefore, by (16),

$$
\operatorname{tr} M_+(i) = \langle\psi_{in}|\,M_+(i)\,|\psi_{in}\rangle.
$$

Since $M_+(i)$'s rank is one, this holds if and only if (36). ∎

Finally, we investigate Example 6.2. Let

$$
M_+ = \sum_{i=1}^d \alpha_i\,|e_i\rangle\langle e_i|, \; M_+ = \sum_{i=1}^d \beta_i\,|e_i\rangle\langle e_i|,
$$

where $\{|e_i\rangle\}_{i=1}^d$ is an orthonormal basis of $\mathcal{H}_{in}$. Below, we assume

$$
\frac{\alpha_1}{\beta_1} > \frac{\alpha_2}{\beta_2} > \cdots > \frac{\alpha_d}{\beta_d}. \tag{37}
$$

The proof of the following lemma is in Appendix B.

24

Lemma 8.3 *Suppose $\alpha_i \geq 0$, $\beta_i > 0$, and (37) holds. Suppose also $\sum_{i=1}^{m} |\gamma_i|^2 > 0$. Then,*

$$\frac{\alpha_1}{\beta_1} = \frac{\sum_{j=1}^{d} |\gamma_j|^2 \alpha_j}{\sum_{j=1}^{d} |\gamma_j|^2 \beta_j}$$

holds if and only if $|\gamma_1| \neq 0$ and

$$\gamma_2 = \gamma_3 = \cdots = \gamma_d = 0.$$

Lemma 8.4 *Suppose unitary matrices $\{U_i\}_{i=1}^{m}$ satisfies*

$$\sum_{i=1}^{m} U_i A U_i^{\dagger} = (\operatorname{ctr} A) \mathbf{1}. \tag{38}$$

Then,

$$\sum_{i=1}^{m} U_i^{\dagger} B U_i = (\operatorname{ctr} B) \mathbf{1}.$$

Proof. By (38), we have

$$c \operatorname{tr} B \operatorname{tr} A = \sum_{i=1}^{m} \operatorname{tr} B U_i A U_i^{\dagger}$$

$$= \sum_{i=1}^{m} \operatorname{tr} U_i^{\dagger} B U_i A.$$

Since this holds for any A, we have the assertion. ∎

Proposition 8.5 *In Example 6.2, suppose $\alpha_i \geq 0$, $\beta_i > 0$. Also suppose (37) holds. Then, $|\psi_{in}\rangle |\psi_R\rangle \succeq^c |\Phi_d\rangle$ is equivalent to the following: there is a surjection $f : \{1, \cdots, m\} \to \{1, \cdots, d\}$, a state vector $|\varphi\rangle$ and unimodular complex numbers ω_i $(i = 1, \cdots, d)$ such that*

$$\left| e_{f(i)} \right\rangle = \omega_i U_i^{\dagger} |\psi_{in}\rangle , \ (i = 1, 2, \cdots, m),$$
$$c = |\{i \ ; \ f(i) = j\}| \ (j = 1, 2, \cdots, d).$$

Proof. For $|\psi_{in}\rangle |\psi_R\rangle \succeq^c |\Phi_d\rangle$ to hold, we have to have

$$\operatorname{tr} |M_+ - sM_-| = \sum_{i=1}^{d} |\alpha_i - s\beta_i|$$

$$= \frac{1}{c} \sum_{i=1}^{m} \left| \langle \psi_{in}| U_i (M_+ - sM_-) U_i^{\dagger} |\psi_{in}\rangle \right|, \ \forall s \geq 0.$$

Therefore, by defining f properly, we have to have

$$\alpha_j = \frac{1}{c} \sum_{i: f(i)=j} \langle \psi_{in}| U_i M_+ U_i^{\dagger} |\psi_{in}\rangle ,$$

$$\beta_j = \frac{1}{c} \sum_{i: f(i)=j} \langle \psi_{in}| U_i M_- U_i^{\dagger} |\psi_{in}\rangle , \ (j = 1, \cdots, d).$$

Thus, if $f(i) = 1$,

$$
\begin{aligned}
\frac{\alpha_1}{\beta_1} &= \frac{\langle \psi_{in} | U_i M_+ U_i^\dagger | \psi_{in} \rangle}{\langle \psi_{in} | U_i M_- U_i^\dagger | \psi_{in} \rangle} \\
&= \frac{\sum_{j=1}^d |\gamma_{i,j}|^2 \alpha_j}{\sum_{j=1}^d |\gamma_{i,j}|^2 \beta_j},
\end{aligned}
$$

where

$$
U_i^\dagger | \psi_{in} \rangle = \sum_{j=1}^d \gamma_{i,j} | e_j \rangle .
$$

By Lemma 8.3, then we should have

$$
\omega_i U_i^\dagger | \psi \rangle = | e_1 \rangle .
$$

Thus,

$$
\alpha_1 = \frac{1}{c} \sum_{i : f(i)=1} \langle \psi_{in} | U_i M_+ U_i^\dagger | \psi_{in} \rangle = \frac{1}{c} |\{ i \, ; \, f(i) = 1 \}| \, \alpha_1,
$$

or

$$
c = |\{ i \, ; \, f(i) = 1 \}| . \tag{39}
$$

Therefore,

$$
\frac{1}{c} \sum_{i \, : \, f(i)=1} U_i^\dagger | \psi_{in} \rangle \langle \psi_{in} | U_i = | e_1 \rangle \langle e_1 | . \tag{40}
$$

Since by Lemma 8.4

$$
\frac{1}{c} \sum_{i=1}^m U_i^\dagger | \psi_{in} \rangle \langle \psi_{in} | U_i = \mathbf{1}
$$

holds, we should have

$$
\frac{1}{c} \sum_{i \, : \, f(i)\neq1} U_i^\dagger | \psi_{in} \rangle \langle \psi_{in} | U_i = \sum_{j=2}^d | e_j \rangle \langle e_j | . \tag{41}
$$

Therefore, with $f(i) > 1$, we should have

$$
\omega_i U_i^\dagger | \psi_{in} \rangle = \sum_{j=2}^d \gamma_j | e_j \rangle ,
$$

and if $f(i) = 2$,

$$
\begin{aligned}
\frac{\alpha_2}{\beta_2} &= \frac{\left| \langle \psi_{in} | U_i M_+ U_i^\dagger | \psi_{in} \rangle \right|}{\left| \langle \psi_{in} | U_i M_- U_i^\dagger | \psi_{in} \rangle \right|} \\
&= \frac{\sum_{j=2}^d |\gamma_j|^2 \alpha_j}{\sum_{j=2}^d |\gamma_j|^2 \beta_j} .
\end{aligned}
$$

26

Then, by Lemma 8.3, we should have

$$\omega_i U_i^\dagger |\psi_{in}\rangle = |e_2\rangle.$$

Therefore, using the same argument as the one derived (39), (40), and (41), we have

$$c = |\{i \,;\, f(i) = 2\}|,$$

$$\frac{1}{c} \sum_{i \,:\, f(i)=2} U_i^\dagger |\psi_{in}\rangle \langle\psi_{in}| U_i = |e_2\rangle \langle e_2|,$$

$$\frac{1}{c} \sum_{i \,:\, f(i)\neq 1,2} U_i^\dagger |\psi_{in}\rangle \langle\psi_{in}| U_i = \sum_{i=3}^{d} |e_i\rangle \langle ei|.$$

Recursively, for each j, we obtain

$$\omega_i U_i^\dagger |\psi_{in}\rangle = \left|e_{\tilde{f}(i)}\right\rangle$$

$$c = |\{i \,;\, f(i) = j\}|$$

$$\frac{1}{c} \sum_{i \,:\, f(i)=j} U_i^\dagger |\psi_{in}\rangle \langle\psi_{in}| U_i = |e_j\rangle \langle e_j|,$$

$$\frac{1}{c} \sum_{i \,:\, f(i)\neq 1,2,\cdots,j} U_i^\dagger |\psi_{in}\rangle \langle\psi_{in}| U_i = \sum_{i=j+1}^{d} |e_i\rangle \langle e_i|.$$

Thus we obtain the assertion of the proposition. ∎

8.3 Entanglement breaking channels which requires entanglement

In [13], they had shown that Bayes error probability of hypothesis testing of a pair of entanglement breaking channel is smaller with a maximally entangled input state than with any separable input states. Likewise, we point out that a maximally entangled state is universally optimal and strictly universally better than any separable state for some families of entanglement breaking channels. Such families of entanglement breaking channels can be composed using Theorem 8.1 and Propositions 8.2 and 8.5.

First, let us compose such family in the form of $\left\{\Lambda_{2,\xi_\theta}^{\mathrm{gp}}\right\}_{\theta\in\Theta}$ using Theorem 8.1. Observe $\Lambda_{2,\xi_\theta}^{\mathrm{gp}}$ is entanglement breaking if and only if $\Lambda_{2,\xi_\theta}^{\mathrm{gp}}\otimes\mathbf{I}\left(|\Phi_2\rangle\langle\Phi_2|\right)$ is separable. By PPT criteria [8], this is equivalent to

$$\xi_\theta^0 + \xi_\theta^3 \geq \left|\xi_\theta^1 - \xi_\theta^2\right|, \ \xi_\theta^1 + \xi_\theta^2 \geq \left|\xi_\theta^0 - \xi_\theta^3\right|, \tag{42}$$

where $\xi_\theta^0 := 1 - \xi_\theta^1 - \xi_\theta^2 - \xi_\theta^3$. If a family $\{\xi_\theta\}_{\theta\in\Theta}$ satisfies does not satisfy the hypothesis of Theorem 8.1 is not true, $\left\{\Lambda_{2,\xi_\theta}^{\mathrm{gp}}\right\}_{\theta\in\Theta}$ is an example of a family of

27

channels with desired properties. In particular, if $\{\xi_\theta\}_{\theta \in \Theta}$ is not on the straight line, this is the case. Even if $\{\xi_\theta\}_{\theta \in \Theta}$ is on a straight line with $\xi_{\theta_1} \neq \xi_{\theta_2}$, if no pair out of $\xi_{\theta_1}^0 / \xi_{\theta_2}^0$, $\xi_{\theta_1}^1 / \xi_{\theta_2}^1$, $\xi_{\theta_1}^2 / \xi_{\theta_2}^2$ and $\xi_{\theta_1}^3 / \xi_{\theta_2}^3$ equals with each other, we also obtain an example of an entanglement breaking channel with desired properties.

Second, consider POVM $\{M_+(i)\}$ such that constituent operators are of unit rank and not orthogonal with each other. Also, define POVM $\{M_-(i)\}$ by (16). Then, by Proposition 8.2, the channel family $\{\Lambda_\theta\}_{\theta \in \{+,-\}}$ defined via (14) has desired property. For example, consider a measurement with POVM

$$M_+(1) = \frac{1}{2a^2} \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}, M_+(2) = \frac{1}{2a^2} \begin{bmatrix} a^2 & -ab \\ -ab & b^2 \end{bmatrix}$$

$$M_+(3) = \frac{1}{2a^2} \begin{bmatrix} 0 & 0 \\ 0 & 2a^2 - 2b^2 \end{bmatrix},$$

$$M_-(i) = \operatorname{tr} M_+(i)\, \mathbf{1} - M_+(i),$$

where $a > b > 0$.

Finally, by Proposition 8.5, we can add another set of examples. Observe

$$\sum_{i,j=0}^{d-1} X_d^i Z_d^j A \left(X_d^i Z_d^j \right)^\dagger = \mathbf{1},$$

where X_d, Z_d are defined by (5). By Proposition 8.5, if there is $|\psi_{in}\rangle |\psi_R\rangle$ with $|\psi_{in}\rangle |\psi_R\rangle \succeq^c |\Phi_d\rangle$, we should have

$$\left| e_{\tilde{f}(i,j)} \right\rangle = \omega'_{ij} \left(X_d^i Z_d^j \right)^\dagger |\psi_{in}\rangle,$$

where $\tilde{f}(i,j)$ is a surjection onto $\{1, \cdots, d\}$ and ω'_{ij} is a unimodular complex number. Therefore, with $f(i', j') = 1$,

$$\left| e_{\tilde{f}(i,j)} \right\rangle = \omega'_{ij} \overline{\omega'_{i'j'}} \left(X_d^i Z_d^j \right)^\dagger X_d^{i'} Z_d^{j'} |e_1\rangle.$$

Therefore, by (6), there is a surjection $f(i,j)$ onto $\{1, \cdots, d\}$ and a unimodular complex number ω_{ij} with

$$\left| e_{f(i,j)} \right\rangle = \omega_{ij} X_d^i Z_d^j |e_1\rangle.$$

For example, let

$$|e_1\rangle = \sum_{i=1}^{d} a_i |i\rangle,$$

where $a_i > 0$ and $a_i \neq a_j$ (i, j). Then, $X_d |e_1\rangle$ is neither parallel or orthogonal to $|e_1\rangle$. Therefore, the conditions indicated by Proposition 8.5 are not satisfied, and we have a channel family with desired property.

9   Iterative use of a channel

Allowed to use given channel Λ_θ for n times, one may send in identical n-copies of an input (*identical repetition*), or create a large entangled state in $\mathcal{H}_{in}^{\otimes n}$ and send to the channels $\Lambda_\theta^{\otimes n}$ (*parallel repetition*), or modify the input depending on the output of the previous use of the channel (*sequential repetition*). By definition, an identical repetition is a special case of a parallel repetition, which, in turn, is a special case of a sequential repetition.

The final output state of the identical repetition with the input state $|\psi\rangle^{\otimes n} \in (\mathcal{H}_{in} \otimes \mathcal{H}_R)^{\otimes n}$ and the parallel repetition with the input state $|\psi^n\rangle \in (\mathcal{H}_{in} \otimes \mathcal{H}_R)^{\otimes n}$ is

$$\rho_{\mathrm{if},\theta}^n = \{\Lambda_\theta \otimes \mathbf{I}\,(|\psi\rangle\,\langle\psi|)\}^{\otimes n} \in \mathcal{B}\left((\mathcal{H}_{out} \otimes \mathcal{H}_R)^{\otimes n}\right),$$

and

$$\rho_{\mathrm{pf},\theta}^n = \Lambda_\theta^{\otimes n} \otimes \mathbf{I}\,(|\psi^n\rangle\,\langle\psi^n|) \in \mathcal{B}\left((\mathcal{H}_{out} \otimes \mathcal{H}_R)^{\otimes n}\right),$$

respectively. To describe the final output state of sequential repetition, we introduce a series of Hilbert spaces $\{\mathcal{H}_{in,i}\}_{i=1}^n$, $\{\mathcal{H}_{out,i}\}_{i=1}^n$, $\mathcal{H}_R^n$, where $\mathcal{H}_{in,i} \simeq \mathcal{H}_{in}$ and $\mathcal{H}_{out,i} \simeq \mathcal{H}_{out}$ ($i = 1, \cdots, n$), and a series of completely positive trace preserving maps $\{\Upsilon_i\}_{i=1}^{n-1}$ from $\mathcal{B}(\mathcal{H}_{out,i} \otimes \mathcal{H}_R^n)$ to $\mathcal{B}(\mathcal{H}_{in,i+1} \otimes \mathcal{H}_R^n)$. Here, dimension of $\mathcal{H}_R^n$ is finite and large enough (in fact, $\dim \mathcal{H}_R^n = (\dim \mathcal{H}_{in})^{n+1}(\dim \mathcal{H}_{out})^n$ is enough.) With the initial state $|\psi\rangle \in \mathcal{H}_{in,1} \otimes \mathcal{H}_R^n$, the final output state of the sequential scheme is

$$\rho_{\mathrm{sf},\theta}^n := (\Lambda_\theta \otimes \mathbf{I}) \circ \Upsilon_{n-1} \cdots \circ (\Lambda_\theta \otimes \mathbf{I}) \circ \Upsilon_2 \circ (\Lambda_\theta \otimes \mathbf{I}) \circ \Upsilon_1 \circ (\Lambda_\theta \otimes \mathbf{I})\,(|\psi\rangle\,\langle\psi|),$$
$$\in \mathcal{B}(\mathcal{H}_{out,n} \otimes \mathcal{H}_R^n)$$

to which the measurement M_n is applied.

Theorem 9.1 *Let $\{\Lambda_\theta\}_{\theta\in\Theta}$ be covariant or contravariant channels. Then, the universally optimal identical repetition strategy achieves the figure of merit that can be achieved by the universally optimal sequential repetition strategy. Here the optimal input state is $|\psi_{opt}\rangle^{\otimes n}$, where $|\psi_{opt}\rangle$ is as of (8).*

Proof. By Proposition 2.2, we only have to compose a CPTP map $\tilde{\Gamma}^n$ with

$$\rho_{\mathrm{sf},\theta}^n = \tilde{\Gamma}^n\left(\{\Lambda_\theta \otimes \mathbf{I}\,(|\psi_{opt}\rangle\,\langle\psi_{opt}|)\}^{\otimes n}\right),$$

where, with $\mathcal{H}_R \simeq \mathcal{H}_{in}$,

$$\{\Lambda_\theta \otimes \mathbf{I}\,(|\psi_{opt}\rangle\,\langle\psi_{opt}|)\}^{\otimes n} \in \mathcal{B}\left((\mathcal{H}_{out} \otimes \mathcal{H}_R)^{\otimes n}\right).$$

The composition of $\tilde{\Gamma}^n$ is as follows. Define $\mathcal{H}'_{in,i}$, $\mathcal{H}'^n_R$ with the same dimension as $\mathcal{H}_{in,i}$, $\mathcal{H}_R^n$ which would have used in the sequential repetition protocol resulting the final state $\rho_{\mathrm{sf},\theta}^n$. Prepare $|\psi\rangle$ in $\mathcal{H}'_{in,1} \otimes \mathcal{H}'^n_R$, and apply Γ, which

is composed in the proof of Theorem 4.7, jointly to $\mathcal{H}'_{in,1}$-part of $|\psi\rangle$ and $\mathcal{H}_R$-part of $\Lambda_\theta \otimes \mathbf{I}(|\psi_{opt}\rangle \langle\psi_{opt}|) \in \mathcal{B}(\mathcal{H}_{out} \otimes \mathcal{H}_R)$, producing $(\Lambda_\theta \otimes \mathbf{I})(|\psi\rangle \langle\psi|)$ in the space $\mathcal{H}_{out} \otimes \mathcal{H}_R^{n'}$. Then apply Υ_1, producing $\Upsilon_1 \circ (\Lambda_\theta \otimes \mathbf{I})(|\psi\rangle \langle\psi|)$ in $\mathcal{H}'_{in,2} \otimes \mathcal{H}_R^{n'}$. Repeating this for n times, composition of $\tilde{\Gamma}^n$ is done. $\blacksquare$

Proposition 9.2 *Consider the family* $\{\Lambda_\theta\}_{\theta\in\{+,-\}}$ *in Examples 6.1 and 6.2. Then, a universally optimal input state for parallel repetition is* $|\Phi_d\rangle^{\otimes n}$ *(identical repetition).*

Proof. If Λ_θ is in the form of (15), or of (15), so is $\Lambda_\theta^{\otimes n}$. Therefore, $|\Phi_{d^n}\rangle = |\Phi_d\rangle^{\otimes n}$ is optimal. $\blacksquare$

This proposition motivates following definition of *classical adaptation*: Given $\Lambda_\theta^{\otimes n}$, we divide this into $\Lambda_\theta^{\otimes n_1}$, $\Lambda_\theta^{\otimes n_2}, \cdots, \Lambda_\theta^{\otimes n_m}$, with $\sum_{i=1}^m n_i = n$. We know that preparing input state separately in each block $|\psi_1\rangle \in \mathcal{H}_{in}^{\otimes n_1}$, $|\psi_2\rangle \in \mathcal{H}_{in}^{\otimes n_1}, \cdots, |\psi_m\rangle \in \mathcal{H}_{in}^{\otimes n_m}$ can achieve the same as the optimal parallel repetition. So the question arises whether we can do better by choosing $\left|\psi_j^{x^{j-1}}\right\rangle$ depending on the data $x^{j-1} = (x_1, x_2, \cdots, x_{j-1})$ from measurements M_1, $M_2^{x^1}$, $\cdots, M_{j-1}^{x^{j-2}}$ applied to $|\psi_1\rangle$, $\left|\psi_2^{x^1}\right\rangle$, $\cdots$, $\left|\psi_{j-1}^{x^{j-2}}\right\rangle$, respectively. (Note here the measurement at jth step is depends on the previous data sequence $x^{j-1} = (x_1, x_2, \cdots, x_{j-1})$.)

Theorem 9.3 *Consider a channel family in Example 6.1 or 6.2. Then, classical adaptation does not improve identical repetition.*

Proof. Let $\vec{M}^j := \left\{M_1, M_2^{x^j}, \cdots, M_j^{x^{j-1}}\right\}_{x^{j-1}}$ and $\vec{\psi}^j := \left\{|\psi_1\rangle, \left|\psi_2^{x^1}\right\rangle, \cdots, \left|\psi_{j-1}^{x^{j-1}}\right\rangle\right\}_{x^{j-1}}$. Also, $p_{\theta,\vec{M}^m,\vec{\psi}^m}(t)$ is the probability of choosing the decision t when sequence of adaptive measurements $\vec{M}^m$ and inputs $\vec{\psi}^m$ are chosen. Then, with the prior distribution $\pi(\theta)$, the minimized risk is

$$\inf_{\vec{M}^m,\vec{\psi}^m} \sum_{\theta,t} \pi(\theta)\, p_{\theta,\vec{M}^m,\vec{\psi}^m}(t)\, l_\theta(t)$$

$$= \inf_{\vec{M}^{m-1},\vec{\psi}^{m-1}} \sum_{x^{m-1}} \inf_{M_m^{x^{m-1}},\rho_{x^{m-1}}} \sum_t \sum_\theta \pi(\theta) \times$$

$$p_{\theta,\vec{M}^{m-1},\vec{\psi}^{m-1}}\left(x^{m-1}\right) \operatorname{tr}\left\{\Lambda^{\otimes n_m} \otimes \mathbf{I}(\rho_{x^{m-1}}) M_m^{x^{m-1}}(t)\right\} l_\theta(t),$$

Let us denote the marginal distribution of x^{m-1} and conditional distribution of θ given x^{m-1} by

$$p_{\pi,\vec{M}^{m-1},\vec{\psi}^{m-1}}\left(x^{m-1}\right) := \sum_\theta \pi(\theta)\, p_{\theta,\vec{M}^{m-1},\vec{\psi}^{m-1}}\left(x^{m-1}\right),$$

$$\tilde{\pi}_{\vec{M}^{m-1},\vec{\psi}^{m-1}}\left(\theta|x^{m-1}\right) := \pi(\theta)\, p_{\theta,\vec{M}^{m-1},\vec{\psi}^{m-1}}\left(x^{m-1}\right) / p_{\pi,\vec{M}^{m-1},\vec{\psi}^{m-1}}\left(x^{m-1}\right),$$

respectively. Then, the minimized risk is

$$
\inf_{\vec{M}^m,\vec{\psi}^m} \sum_{\theta,t} \pi\left(\theta\right) p_{\theta,\vec{M}^m,\vec{\psi}^m}\left(t\right) l_\theta\left(t\right)
$$

$$
= \inf_{\vec{M}^{m-1},\vec{\psi}^{m-1}} \sum_{x^{m-1}} p_{\pi,\vec{M}^{m-1},\vec{\psi}^{m-1}}\left(x^{m-1}\right) \times
$$

$$
\inf_{M_m^{x^{m-1}},\rho_{x^{m-1}}} \sum_{\theta,t} \tilde{\pi}_{\vec{M}^{m-1},\vec{\psi}^{m-1}}\left(\theta|x^{m-1}\right) \mathrm{tr}\left\{\Lambda^{\otimes n_m}\otimes\mathbf{I}\left(\rho_{x^{m-1}}\right)M_m^{x^{m-1}}\left(t\right)\right\} l_\theta\left(t\right).
$$

By definition of a universally optimal state, infimum over $\rho_{x^{m-1}}$ can be achieved by $\rho_{x^{m-1}} = |\Phi_d\rangle\langle\Phi_d|^{\otimes n_m}$, which does not depends on the data sequence x^{m-1}. Therefore, we can merge the last two steps into one; depending on x^{m-2}, we send $\rho_{x^{m-2}}\otimes|\Phi_d\rangle\langle\Phi_d|^{\otimes n_m}$ into $\Lambda^{\otimes n_{m-1}+n_m}\otimes\mathbf{I}$ and apply $M_{m-1}^{x^{m-2}}$ and $M_m^{x^{m-1}}$, successively. Repeating this process, we can get rid of classical adaptation. ■

References

[1] P. M. Alberti, A. Uhlmann, "Stochastic Linear Maps and Transition Probability", Lett. Math. Phys. 7 107-112 (1983)

[2] F. Buscemi, "Comparison of Quantum Statistical Models: Equivalent Conditions for Sufficiency", Communications in Mathematical Physics Vol. 310, No. 3, 625-647 (2012)

[3] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Optimal estimation of group transformations using entanglement, Phys. Rev. A 72 042338 (2005)

[4] A. Fujiwara, "Estimation of SU(2) operation and dense coding: An information geometric approach," Phys. Rev. A, vol. 65, 012316 (2002)

[5] A. Fujiwara and Hiroshi Imai, "Quantum parameter estimation of a generalized Pauli channel," J. Phys. A: Math. Gen., vol. 36, pp. 8093-8103 (2003)

[6] A. Fujiwara, "Estimation of a generalized amplitude-damping channel," Phys. Rev. A, vol. 70, 012317 (2004)

[7] A. Fujiwara and H. Imai, "A fibre bundle over manifolds of quantum channels and its application to quantum statistics," J. Phys. A: Math. Theor., vol. 41, 255304 (2008)

[8] M. Horodecki, P. Horodecki and R. Horodecki, "Separability of mixed states: necessary and sufficient conditions", Phys. Lett. A223, 8 (1996)

[9] M. Keyl, R. Werner, "Optimal Cloning of Pure States, Judging Single Clones", J. Math. Phys. 40, 3283(1999)

[10] K. Matsumoto, "A quantum version of randomization criteria" (2010)

[11] M. F. Sacchi, "Optimal discrimination of quantum operations", Phys. Rev. A 71, 062340 (2005)

[12] M. F. Sacchi, "Minimum error discrimination of Pauli channels", J. Opt. B 7, S333 (2005).

[13] M. F. Sacchi, "Entanglement can enhance the distinguishability of entanglement-breaking channels", Phys. Rev. A 72, 014305 (2005)

[14] R. Werner, "Optimal Cloning of Pure States", Phys. Rev. A 58, 1827 (1998)

[15] M. Ziman, "Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments", Phys. Rev. A 77, 062112 (2008)

[16] M. Ziman, "Single-shot discrimination of quantum unitary processes", Journal of Modern Optics, Vol. 57, No. 3, pp. 253-259 (2010).

A Proof of Lemma 7.4

Lemma A.1 *Suppose $x_i \geq 0$ $(i = 1, \cdots, d)$ Then, $Ang(x) \neq \emptyset$ if and only if*

$$x_1^\downarrow \leq \sum_{i=2}^{d} x_i^\downarrow. \tag{43}$$

Proof. Obviously, we only have to prove 'if'. If $d = 3$, the assertion follows from triangle inequality. Suppose the assertion is true for $d - 1$, or for any $y = (y_1, y_2, \cdots, y_{d-1})$ with

$$y_1 \geq y_2 \geq \cdots \geq y_{d-1}$$

and

$$y_1 \leq \sum_{i=2}^{d-1} y_i,$$

we have $Ang(y) \neq \emptyset$. Suppose $x_{d-1}^\downarrow + x_d^\downarrow \leq x_1^\downarrow$ and

$$x_1^\downarrow \leq \sum_{i=2}^{d} x_i^\downarrow = \sum_{i=2}^{d-2} x_i^\downarrow + x_{d-1}^\downarrow + x_d^\downarrow.$$

hold. Then,

$$y_1 := x_1,\ y_2 := x_2, \cdots, y_{d-2} := x_{d-2},\ y_{d-1} = x_{d-1}^\downarrow + x_d^\downarrow,$$

$Ang(y) \neq \emptyset$ by the hypothesis of induction. Therefore, $Ang(x) \neq \emptyset$ holds. On the other hand, suppose $x_{d-1}^\downarrow + x_d^\downarrow > x_1^\downarrow$. Observe $x_1^\downarrow \geq x_2^\downarrow \geq \cdots \geq x_d^\downarrow \geq 0$ yields

$$\sum_{i=1}^{d-2} x_i^\downarrow \geq x_{d-1}^\downarrow + x_d^\downarrow.$$

32

Therefore, with

$$y_1 := x_{d-1}^{\downarrow} + x_d^{\downarrow}, \; y_2 = x_1^{\downarrow}, \; y_3 = x_2^{\downarrow}, \cdots, y_{d-1} = x_{d-2}^{\downarrow},$$

by the hypothesis of induction, $Ang\,(y) \neq \emptyset$ holds. Therefore, $Ang\,(x) \neq \emptyset$ holds. After all, we have the assertion. ∎

Lemma A.2 *Suppose $d = 3$ and $x_i \geq 0$, $x_1 \geq x_2 \geq x_3$. Then, if $x_1 < x_2 + x_3$,*

$$Ang\,(x) = \{(\omega_1, \omega_2), (\overline{\omega_1}, \overline{\omega_2})\}.$$

If $x_1 = x_2 + x_3$,

$$Ang\,(x) = \{(-1, 1, 1)\}.$$

Lemma A.3 *Suppose $d \geq 4$, $x_1 \geq x_2 \geq \cdots \geq x_d > 0$, and*

$$x_1 < \sum_{i=2}^{d} x_i. \tag{44}$$

Then, $Ang\,(x)$ is a $(d-3)$-dimensional smooth manifold.

Proof. Let

$$z_k := \sum_{i=k}^{d-1} x_i \omega_i + x_d, \quad (k = 3, \cdots, d-1),$$
$$z_d := x_d, \quad r_k := |z_k|,$$
$$\vec{r} := (r_3, r_4, \cdots, r_{d-1}).$$

Suppose $\vec{r}$ is fixed. Then, length of each edge of each triangle $z_1 z_{k-1} z_k$ ($k = 3, \cdots, d$) is decided, and $\vec{\omega}$ can take only finite possible values. Also, the map from $\vec{r}$ to $\vec{\omega}$ is smooth. Therefore, we use $\vec{r}$ as a local coordinate of $Ang\,(x)$. Let $A\,(x)$ be the set of all $\vec{r}$s such that $\vec{\omega} \in Ang\,(x)$. Below, we show the interior $A\,(x)^{\circ}$ of $A\,(x)$ is non-empty. Then, the assertion of the lemma immediately follows.

An element of $A\,(x)$ is constructed as follows. We first fix ω_{d-1}, r_{d-1}, then ω_{d-2}, r_{d-2}, ..., ω_{k+1}, r_{k+1}. We choose r_k so that the following (46) and (47) are satisfied (then, ω_k can take only one of two possible values.); First, by

$$r_k = |z_{k+1} + x_k \omega_k|, \tag{45}$$

existence of ω_k is equivalent to

$$|r_{k+1} - x_k| \leq r_k \leq r_{k+1} + x_k. \tag{46}$$

Also, for $\omega_{k-1}, \cdots, \omega_1$ to exist, by Lemma A.1, it is necessary and sufficient that

$$x_1 - \sum_{i=2}^{k-1} x_i \leq r_k \leq \sum_{i=1}^{k-1} x_i. \tag{47}$$

33

Therefore, $A(x)$ is the set of $\vec{r}$s with (46) and (47) for each $k = 3, \cdots, d-1$. Therefore, $A(x)^\circ$ is the set of all $\vec{r}$s with

$$|r_{k+1} - x_k| < r_k < r_{k+1} + x_k, \tag{48}$$

and

$$x_1 - \sum_{i=2}^{k-1} x_i < r_k < \sum_{i=1}^{k-1} x_i \tag{49}$$

for each $k = 3, \cdots, d-1$.

This $A(x)^\circ$ is non-empty due to the following reasons. By (44), we have

$$x_1 - \sum_{i=2}^{d-2} x_i < r_d + x_{d-1}.$$

Also, by $x_{d-2} \geq x_{d-1} \geq r_d > 0$, we have

$$|r_d - x_{d-1}| < \max\{r_d, x_{d-1}\} \leq x_{d-2} < \sum_{i=1}^{d-2} x_i.$$

Therefore, combining these, the overlap of the set

$$\left\{ r_{d-1} \, ; \, x_1 - \sum_{i=2}^{d-2} x_i < r_{d-1} < \sum_{i=1}^{d-2} x_i \right\},$$

and the set

$$\{ r_{d-1} \, ; \, |r_d - x_{d-1}| < r_{d-1} < r_d + x_{d-1} \}$$

is not empty. Recursively, suppose r_k with (48) and (49) exists. Then, by (48),

$$r_k - x_{k-1} < r_k < \sum_{i=1}^{k-2} x_i, \;\; x_1 - \sum_{i=2}^{k-2} x_i < r_k + x_{k-1}.$$

Also, by $x_1 \geq x_2 \geq \cdots \geq x_d$ and $x_i > 0$,

$$x_{k-1} - r_k \leq x_{k-1} < \sum_{i=1}^{k-2} x_i \, .$$

Therefore,

$$|r_k - x_{k-1}| < \sum_{i=1}^{k-2} x_i, \; x_1 - \sum_{i=2}^{k-2} x_i < r_k + x_{k-1}.$$

Therefore, there is r_{k-1} with

$$|r_k - x_{k-1}| < r_{k-1} < r_k + x_{k-1}$$

and

$$x_1 - \sum_{i=2}^{k-2} x_i < r_k < \sum_{i=1}^{k-2} x_i.$$

Therefore, there exists $\vec{r}$ such that (48) and (49) hold for each k, or equivalently, $A(x)^\circ$ is non-empty. ∎

Lemma A.4 *Suppose $d \geq 4$ and $x_i > 0$ $(i = 1, \cdots, d)$. Then $Ang(x)$ contains an element $\vec{\omega}$ such that the set $\{\omega_1, \omega_2, \cdots, \omega_{d-1}\}$ contains at least three distinct elements.*

Proof. Suppose ω_1, ω_2, $\cdots$, ω_{d-1} can take at most two distinct values for any element of $Ang(x)$. Let I be a subset of $\{1, \cdots, d-1\}$, and $\omega_i = \nu_I$ ($i \in I$), $\omega_i = \nu'_I$ ($i \in I^c$). Then, (ν_I, ν'_I) is decided by Lemma A.2. Moving I over all the subsets of $\{1, \cdots, d-1\}$, (ν_I, ν'_I) can move over discretely many values. This contradicts with Lemma A.3. ∎

Now, we are in the position to state the proof of Lemma 7.4.
Proof. (Lemma 7.4) When $d = 3$, the assertion is trivial. So suppose $d \geq 4$. Let

$$\vec{\omega}(t) = \left(e^{\sqrt{-1}\eta_1(t)}, e^{\sqrt{-1}\eta_2(t)}, \cdots, e^{\sqrt{-1}\eta_{d-1}(t)} \right) \in Ang(x) \subset Ang(x'),$$

where $\eta_i(t)$ are smooth functions. (Such smooth parameter t exists due to Lemma A.3.) Then,

$$\sum_{i=1}^{d-1} x_i e^{\sqrt{-1}\eta_i} + x_d = \sum_{i=1}^{d-1} x'_i e^{\sqrt{-1}\eta_i} + x'_d = 0.$$

Differentiating by t,

$$\sum_{i=1}^{d-1} x_i \dot{\eta}_i e^{\sqrt{-1}\eta_i} = \sum_{i=1}^{d-1} x'_i \dot{\eta}_i e^{\sqrt{-1}\eta_i} = 0. \tag{50}$$

Due to Lemma A.3, with

$$\tilde{\eta} := \left(\dot{\eta}_1, \dot{\eta}_2, \cdots, \dot{\eta}_{d-1} \right),$$

span $\{\tilde{\eta}; (50) \text{ holds}\}$ is $d - 3$ dimensional. Therefore, its orthogonal complement in $\mathbb{R}^{d-1}$ is at most two dimensional. By (50),

$$(x_1 \cos \eta_1, \cdots, x_{d-1} \cos \eta_{d-1}), \; (x_1 \sin \eta_1, \cdots, x_{d-1} \sin \eta_{d-1}),$$
$$(x'_1 \cos \eta_1, \cdots, x'_{d-1} \cos \eta_{d-1}), \; (x'_1 \sin \eta_1, \cdots, x'_{d-1} \sin \eta_{d-1}),$$

are orthogonal to span $\{\tilde{\eta}; (50) \text{ holds}\}$. By Lemma A.4, we can choose η_i so that the set $\{\eta_1, \eta_2, \cdots, \eta_{d-1}\}$ contains at least three distinct values. Therefore, $(x_1 \cos \eta_1, \cdots, x_{d-1} \cos \eta_{d-1})$ and $(x_1 \sin \eta_1, \cdots, x_{d-1} \sin \eta_{d-1})$ are linearly independent, thus can be chosen as a basis of orthogonal complement of span $\{\tilde{\eta}; (50) \text{ holds}\}$. Therefore, there are $a_1, \cdots, a_4$ with

$$(x'_1 \cos \eta_1, \cdots, x'_{d-1} \cos \eta_{d-1}) = a_1 (x_1 \cos \eta_1, \cdots, x_{d-1} \cos \eta_{d-1}) + a_2 (x_1 \sin \eta_1, \cdots, x_{d-1} \sin \eta_{d-1}),$$
$$(x'_1 \sin \eta_1, \cdots, x'_{d-1} \sin \eta_{d-1}) = a_3 (x_1 \cos \eta_1, \cdots, x_{d-1} \cos \eta_{d-1}) + a_4 (x_1 \sin \eta_1, \cdots, x_{d-1} \sin \eta_{d-1}).$$

Therefore,

$$\left(a_1 \cos \eta_i \sin \eta_i + a_2 \sin^2 \eta_i - a_3 \cos^2 \eta_i - a_4 \cos \eta_i \sin \eta_i \right) x_i$$
$$= \left(\frac{a_1 - a_4}{2} \sin 2\eta_i - \frac{a_2 + a_3}{2} \cos 2\eta_i + \frac{a_2 - a_3}{2} \right) x_i = 0.$$

Since $x_i > 0$, we have

$$(a_1 - a_4) \sin 2\eta_i - (a_2 + a_3) \cos 2\eta_i + a_2 - a_3 = 0.$$

Therefore, $a_1 - a_4 = a_2 + a_3 = 0$, since the set $\{\eta_1, \eta_2, \cdots, \eta_{d-1}\}$ contains at least three distinct values by Lemma A.4. Therefore,

$$a_1 - a_4 = a_2 + a_3 = a_2 - a_3 = 0,$$

which means

$$(\cos \eta_1 \, x_1', \cdots, \cos \eta_{d-1} \, x_{d-1}') = a_1 (\cos \eta_1 \, x_1, \cdots, \cos \eta_{d-1} \, x_{d-1}),$$
$$(\sin \eta_1 \, x_1', \cdots, \sin \eta_{d-1} \, x_{d-1}') = a_1 (\sin \eta_1 \, x_1, \cdots, \sin \eta_{d-1} \, x_{d-1}).$$

Since one of $\cos \eta_i$ and $\sin \eta_i$ is always non-zero, we have the assertion. ∎

B   Proof of Lemma 8.3

Proof. (Lemma 8.3)

Observe

$$\frac{\alpha_{d-1}}{\beta_{d-1}} - \frac{\sum_{i=d-1}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=d-1}^{d} |c\gamma|^2 \beta_i} = \frac{|\gamma_d|^2 \beta_d}{\sum_{i=d-1}^{d} |\gamma_i|^2 \beta_i} \left(\frac{\alpha_{d-1}}{\beta_{d-1}} - \frac{\alpha_d}{\beta_d} \right).$$

Therefore, if $\sum_{i=d-1}^{d} |\gamma_i|^2 \neq 0$, we have

$$\frac{\alpha_{d-1}}{\beta_{d-1}} \geq \frac{\sum_{i=d-1}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=d-1}^{d} |\gamma_i|^2 \beta_i}.$$

Next, observe

$$\frac{\alpha_{d-2}}{\beta_{d-2}} - \frac{\sum_{i=d-2}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=d-2}^{d} |\gamma_i|^2 \beta_i} = \frac{\sum_{i=d-1}^{d} |\gamma_i|^2 \beta_i}{\sum_{i=d-2}^{d} |\gamma_i|^2 \beta_i} \left(\frac{\alpha_{d-2}}{\beta_{d-2}} - \frac{\sum_{i=d-1}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=d-1}^{d} |\gamma_i|^2 \beta_i} \right)$$
$$\geq \frac{\sum_{i=d-1}^{d} |\gamma_i|^2 \beta_i}{\sum_{i=d-2}^{d} |\gamma_i|^2 \beta_i} \left(\frac{\alpha_{d-2}}{\beta_{d-2}} - \frac{\alpha_{d-1}}{\beta_{d-1}} \right)$$

Therefore, if $\sum_{i=d-2}^{d} |\gamma_i|^2 \neq 0$, we have

$$\frac{\alpha_{d-2}}{\beta_{d-2}} \geq \frac{\sum_{i=d-2}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=d-2}^{d} |\gamma_i|^2 \beta_i}.$$

Recursively, if $\sum_{i=2}^{d} |\gamma_i|^2 \neq 0$, we have

$$\frac{\alpha_2}{\beta_2} \geq \frac{\sum_{i=2}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=2}^{d} |\gamma_i|^2 \beta_i}.$$

Observe

$$\frac{\alpha_1}{\beta_1} - \frac{\sum_{i=1}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=1}^{d} |\gamma_i|^2 \beta_i}$$

$$= \frac{\sum_{i=2}^{d} |\gamma_i|^2 \beta_i}{\sum_{i=1}^{d} |\gamma_i|^2 \beta_i} \left(\frac{\alpha_1}{\beta_1} - \frac{\sum_{i=2}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=2}^{d} |\gamma_i|^2 \beta_i} \right)$$

$$\geq \frac{\sum_{i=2}^{d} |\gamma_i|^2 \beta_i}{\sum_{i=1}^{d} |\gamma_i|^2 \beta_i} \left(\frac{\alpha_1}{\beta_1} - \frac{\alpha_2}{\beta_2} \right)$$

Therefore, due to (37),

$$\frac{\alpha_1}{\beta_1} = \frac{\sum_{i=1}^{d} |\gamma_i|^2 \alpha_i}{\sum_{i=1}^{d} |\gamma_i|^2 \beta_i}$$

implies $\sum_{i=2}^{d} |\gamma_i|^2 = 0$. Thus we have the assertion. ∎